

T.C.
MİLLÎ EĞİTİM BAKANLIĞI

MODÜL 2

ÖLÇME VE DEĞERLENDİRME

YAZARLAR
Doç. Dr. Güçlü ŞEKERCİOĞLU
Dr. Hayri Eren SUNA

TEMMUZ 2022
ANKARA



1. ÖLÇME VE DEĞERLENDİRMENİN TEMEL KAVRAMLARI

Doç. Dr. Güçlü ŞEKERCİOĞLU

1.1. EĞİTİM SİSTEMİNDE ÖLÇME VE DEĞERLENDİRMENİN YERİ: Günlük yaşamda "eğitim sistemi" terimini sık kullanırız. Eğitimin gerçekten bir sistem olup olmadığını değerlendirelim ancak öncelikle sistemin ne olduğuna bir bakalım: Sistem, en az bir hedefi gerçekleştirmek üzere uygun ve değişik ögelerden oluşan dirik[1] bir örüntüdür[2]. Bir bütüne sistem diyebilmek için girdi, süreç ve çıktı ögelerinin olması gerekir. Eğitim de bir sistemdir çünkü birincisi gerçekleştirmek istediği hedefler bulunmaktadır. Hedefler şu biçimde tanımlanabilir:

1. ***Uzak hedefler:*** Devletlerin hedefleridir, bazı ülkelerde çerçevesi anayasa ile güvenceye alınmıştır.
2. ***Genel hedefler:*** Daha işevuruk hedeflerdir. Kurumların (MEB, MEB genel müdürlükleri, il millî eğitim müdürlükleri, okul yönetimleri vb.) hedefleri vb.
3. ***Özel hedefler:*** Dersin hedefleridir, bir dersin kazanımlarını kapsar.

Eğitimin bir sistem olmasının ikinci nedeni, bu hedefleri gerçekleştirmek için farklı ögelerden oluşmasıdır:

1. ***Girdi:*** Eğitim-öğretim için gerekli her şey eğitim sisteminin girdisini oluşturur. Örneğin *öğrenci, öğretmen, okul, çalışanlar, yönetim, eğitim programı vb.*
2. ***Süreç:*** Girdilerin hedefler doğrultusunda bir araya getirildiği, dersin hedeflerinin öğrenciye kazandırıldığı eğitim durumlarını kapsar. Süreç, öğretmenler, okul yönetimi, il yöneticileri ülke genelinde belirli aralıklarla kontrol edilmektedir. Kontrollerin sonucu olumlu ise süreç devam etmekte değilse sürece müdahale edilmektedir.
3. ***Çıktı:*** Öğrencilerin sahip olduğu niteliklerdir. Eğitim sistemi için çıktılar, dar anlamda öğrencinin bir dersin sonunda ulaştığı hedefler ve kazanımları ifade ederken geniş anlamda öğrencinin öğretim sürecinin başlangıcından sonuna kadar sahip olduğu bilgi beceri ve yetkinlik anlarını ifade etmektedir. Yönetimler girdiyi kontrol ederler. Eğitim sisteminde öğrenci hazırbulunuşlukları da dar anlamda eğitim sisteminin girdilerini oluşturur ve öğretmenler hazırbulunuşluk girdisini ölçerek kontrol etmektedirler. Çıktılar hedeflerle tutarlı mı diye kontrol edilmektedir. Çıktılar hedeflerle tutarlı ise eğitim sistemi doğru bir şekilde işliyor demektir. Eğer çıktılar hedeflerle tutarlı değilse eğitim sisteminde bir şeyler aksıyor demektir.

Eğitimde kontrol (denetleme) ölçme ve değerlendirme ile mümkündür. Gözlenmeyen süreçler ve ürün değerlendirilemez. Diğer taraftan içinde bulunduğumuz yüzyıl ile birlikte eğitimde ve dolayısıyla ölçme ve değerlendirmede bir paradigma değişimi söz konusudur.

1.2. YİRMİ BİRİNCİ YÜZYIL DEĞİŞEN EĞİTİM PARADİGMASI:

20. yüzyıla hâkim olan kültürel iklim, sanayileşme idi ve dolasıyla eğitim bu iklimle uyumlu biçimde şekillenmişti. Özellikle ABD'de 1930-1950'lerde okullardan "*eğitim fabrikaları*" olarak söz ediliyor, eğitim de bu anlayışa uygun biçimde kurgulanıyordu. Ülkemizde de o yıllarda benzer bir anlayışın olduğunu söylemek olanaklı. Atatürk'ün ortaya koyduğu «*muasır medeniyetler seviyesine ulaşmak*» Türkiye'nin hedefi ve bir nevi eğitimin de uzak hedefi oldu. Eğitimin genel hedefi de pek çok ülkede kabaca *"toplumun bireylerden beklediği görev ve sorumluluklar doğrultusunda gerekli bilgi ve donanıma sahip olma"* olarak tanımlandı.

---

[1] Canlılığını sürdüren, sürekli kendisini yenileyen ve gelişen.

[2] Ögelerden oluşan ve aralarında ilişki bulunan bir bütün.



Geçtiğimiz yüzyılın özellikle son çeyreğinde bilgi ve iletişim teknolojisinin öneminin artması toplumlarda, üretimde ve eğitimde bir paradigma değişimine yol açtı. Artık sanayi toplumlarından bilgi toplumlarına hızlı bir dönüşüm baş gösterdi. Değişen tanımlamalar ve kavramlar Tablo 1'de verilmiştir.

Tablo 1. *Değişen üretim paradigması*

Sanayi Paradigması	**Bilgi ve İletişim Teknolojisi Paradigması**
Enerji yoğun	Bilgi yoğun
Çizim bürolarında yapılan tasarım ve mühendislik	Bilgisayar destekli tasarım
Ardışık tasarım ve üretim	Eş zamanlı mühendislik
Standart tasarım	Sipariş üzerine tasarım
Tahsis edilmiş tesis donanım	Esnek üretim sistemleri
Otomasyon	Sistemasyon
Tek firma	Ağ yapı
Hiyerarşik yapılar	Düz yapılar
Bölünmüş	Bütünleşik
Servisi olan üretim	Ürünleri olan servis
Merkezî yapı	Dağılmış yapı
Özgün beceri	Çoklu beceri
Hükûmet kontrolü	Hükûmetin bilgi, eş güdüm ve üretimi yönlendirmesi
Planlama	Vizyon

Üretimde yaşanan bu değişim eğitim-öğretim ortamlarına da yansımaya başladı. Değişen tanımlamalar ve kavramlar Tablo 2'de verilmiştir.

Tablo 2. *Değişen üretim paradigması*

Sanayi Paradigması	**Bilgi ve İletişim Teknolojisi Paradigması**
Sınıflarda eğitim	Bireysel araştırma
Pasif öğrenme	Yaparak-yaşayarak öğrenme
Bireysel çalışma	Grupla çalışma
Her şeyi bilen öğretmen	Rehberlik eden öğretmen
Sabit içerik	Esnek içerik
Homojen	Heterojen

Diğer taraftan beşerî bilimler ve özelinde eğitim bilimlerinde bilişsel, duyuşsal ve devinimsel (psikomotor) olarak üç boyutta tanımlanan insan davranışları, birbirinden ayrı kümelermiş gibi ele alındı, araştırıldı, hakkında bilgi üretilmeye çalışıldı. Hâlbuki insan bu parçaların toplamından daha anlamlı bir varlıktı.

20. yüzyılın son çeyreğinde ve özellikle 21. yüzyılda ekonomi çevreleri ve ekonomistler eğitime yoğun bir ilgi göstermeye başladı. Makro ekonomiyi ve gelecek ekonomi perspektiflerini ele alan değerlendirmelerde eğitime vurgu yapmayan ekonomist pek kalmadı. Ekonomistler son yıllarda eğitime neden daha çok vurgu yapıyorlar? Araştırmalar, paradigma değişimine uygun insan gücü kaynağının ancak nitelikli eğitimle gerçekleşebileceğini ortaya koymakta. Niteliğin ölçülmesi çok zor ancak nicel veriler açısından değerlendirdiğimizde bir ülkede ortalama eğitim bir birim arttığında, ekonomik büyüme %0,7 artmakta. Bu büyümenin yarısı doğrudan eğitimin katkısı iken diğer yarısı eğitim ve ekonomi arasındaki etkileşimden



kaynaklanmakta. Dikkat edilirse PISA, TIMSS ve PIRLS gibi dünyanın en geniş çaplı uluslararası eğitim araştırmalarını yapan örgütler ya doğrudan ekonomi ya da yine siyasi-ekonomik politikaların üretildiği örgütler.

Dünya Ekonomik Forumuna göre 2025 yılında çalışanların sahip olması gereken 10 temel beceri:

- Analitik düşünme ve yenilik/yaratıcılık
- Etkin öğrenme ve öğrenme stratejileri
- Karmaşık problem çözme
- Eleştirel düşünme ve analiz
- Yaratıcılık, orijinallik
- Liderlik ve sosyal etki
- Teknoloji kullanımı

Bu değişim doğrultusunda dünyanın biz eğitimcilerden beklediği görev ve sorumluluklar, eğitim ve öğretim ortamlarını yeniden tasarlamamıza ve denetleme biçimlerimizi farklılaştırmamıza neden oldu; bu değişim devam ediyor.

Eğitimde denetleme (kontrol) ölçme ve değerlendirme ile olanaklıdır ancak değişen hedeflerde bireylere kazandırılacak olan yeni davranışların ölçülmesi ve değerlendirilmesinde de doğal olarak bir paradigma değişimi söz konusu. Geleneksel ölçme anlayışının ne yazık ki bu paradigma değişimine pek uygun olmadığı söylenebilir. Bireylerden beklenen becerilerin karmaşıklığı arttıkça ölçme modellerini revize etme gereksinimi ortaya çıkmakta. Bir ölçme işleminde yalnızca "görünce tanıma, sorulunca söyleme" gibi davranışların ölçülmesi bizi yeni hedeflere ne kadar ulaştırabilir? 20. yüzyıl eğitim ve ölçme zihniyeti ile 21. yüzyıl becerileri kazandırılamaz.

1.3. ÖLÇME, ÖLÇÜT VE DEĞERLENDİRMENİN KAVRAMSAL TEMELLERİ:

1.3.1. Ölçme: Öğrencinin belirli bir özelliğini[3] gözleyerek[4] o özelliğe sayı, sembol ya da sıfat/kategori adı[5] verme işlemidir. Örneğin

- *Kerim Temel Yeterlilik Testinden (TYT) 450 puan aldı.*
- *Can'ın genel yetenek testi sonucu, zekâ bölümünün 132 olduğu belirlendi.*
- *Zeynep yapılan test sonucunda 1. oldu.*

Ölçmede gözlenen özelliğin bir betimlemesi yapılır. Bu nedenle tek başına bir anlam ifade etmez. Ölçmenin anlam ifade etmesi değerlendirme ile olanaklıdır. Diğer taraftan ölçmeye temel olan durum "*fark*"tır. Bir anlamda ölçmenin farktan doğduğu ifade edilebilir. Fark, bireyler arasında olabileceği gibi bireylerin farklı zaman dilimlerinde ortaya koydukları özellikler arasında da olabilir.

Eğitimde ölçme dediğimizde aklımıza ilk gelen genellikle sınavlardır. Eğitimde ölçme yalnızca sınavlardan ibaret değildir. Eğitimde akademik başarı dışında yetenek, ilgi, tutum, özel gereksinim gereken alanlar vb. özellikler de ölçmeye konu olur.

Bir eğitim sisteminde ölçme ve değerlendirme olmaksızın eğitim sisteminin geçerli ve güvenilir kontrolü sağlanamaz. Bu kontrol, sistemin tüm ögeleri için geçerlidir.

Ölçülmeyen, kontrol edilmeyen süreçler denetlenemez ve denetlenemeyen süreçler ise değerlendirilemez. Ölçme yapılmadan sisteme yönelik girdi, süreç ve çıktılar kontrol altına

[3] Eğitimde ölçmeye konu olan genellikle zekâ, başarı, yetenek, kişilik, tutum vb. psikolojik özelliklerdir.
[4] Öğrenciyi yapay uyarıcılara maruz bırakarak gerçekleştirilen yapılandırılmış bir gözlem.
[5] Ölçüm=Ölçme Sonucu. *Örneğin 60, CC ya da orta vb.*



alınamaz. Dolayısıyla bir eğitim siteminde ölçme olmadan denetleme ve kontrol mekanizmasını işletmek mümkün değildir.

1.3.2. Ölçüt: Ölçülen özellik hakkında karar alabilmek / yargıya varabilmek / değerlendirme yapabilmek için dayanak alınan referans noktası[6] ya da referans aralığıdır[7].

1.3.3. Değerlendirme: Bir ölçme sonucunu (ölçüm) en az bir ölçüte vurarak ölçülen nitelik hakkında karar verme / yargıda bulunma işlemidir. Değerlendirmenin basamakları a. Ölçme ⇒ Ölçme sonucu (ölçüm), b. Ölçüt ve c. Karardır. Örneğin

- *Kerim gıda mühendisliği programına yerleştirilmiştir / yerleştirilememiştir.*
- *Can üstün yetenekliler sınıfına kabul edilmiştir / edilmemiştir.*
- *Zeynep dersten geçmiştir / kalmıştır.*
- *Mehmet öğretmen olarak atanmıştır / atanamamıştır.*
- *Hakan burs almaya hak kazanmıştır / burs alamamıştır.*

Ölçüt, ölçme ve değerlendirme arasında köprü görevi görür. Ölçüt değişirse değerlendirme de değişir. Örneğin

Ölçme	Ölçüt	Değerlendirme
*Irmak Yabancı Dil Bilgisi Seviye Tespit Sınavından (YDS) **50** almıştır.*	***50***	***Yüksek lisansa başvuru yapabilir.***
*Irmak Yabancı Dil Bilgisi Seviye Tespit Sınavından (YDS) **50** almıştır.*	***55***	***Doktoraya başvuru yapamaz.***

1.4. ÖLÇME TÜRLERİ: Eğitimde ölçme doğrudan, dolaylı ve türetilmiş olmak üzere üçe ayrılır.

1.4.1. Doğrudan Ölçme (Temel Ölçme): Ölçmeye konu olan özelliğin doğrudan gözlenerek ölçümün elde edildiği ölçme türüdür. Genellikle beş duyu organı ile algılanan özellikler doğrudan gözlenebilirler. Örneğin *bir sınıftaki öğrenci sayısı, parktaki salıncak sayısı, öğrencilerin cinsiyetleri, boyları, kalem tutma becerisi, pas verme becerisi vb.* özellikler doğrudan gözlenebilir.

1.4.2. Dolaylı Ölçme (Göstergeyle Ölçme): Bazı özellikler doğrudan gözlenemez. Bu özellikler ancak onun göstergesi olduğu bilinen ya da kabul edilen davranışlar aracılığı ile gözlenerek ölçülebilir. Örneğin *öğrencilerin dört işlem becerisi, derse yönelik tutumları, genel muhakeme düzeyleri, öz düzenleme becerileri, sınav kaygıları, motivasyon düzeyleri vb.* özellikler ancak dolaylı olarak gözlenebilir.

Eğitimde ölçmeye konu olan özelliklerin pek azı doğrudan ölçülebilir. Söz gelimi yanlışsız okuyabilme, diksiyon, pek çok psikomotor davranış doğrudan gözlenebilir. Diğer taraftan bir öğretmen adayına *"Sınıfta ……. durumuyla karşılaştığınızda nasıl davranırsınız?"* diye sorduğunuzda vereceği sözlü ya da yazılı tepkinin gelecekte nasıl davranacağının bir göstergesi olduğu kabul edilir. Benzer biçimde *"iyi yurttaş"* olma, doğrudan gözlenebilir bir özellik değilken iyi yurttaş olmanın göstergesi olarak kabul edilen *"vergi verme, seçme ve seçilme hakkını kullanma, askerlik vb. görevleri yapma"* gibi davranışları sergileyen kişi "iyi

[6] Örneğin 60 alan geçer vb.

[7] Örneğin 90-109 arası normal zekâdır, DC alan şartlı geçer (aritmetik ortalama – ½ standart sapma) vb.



yurttaş" olarak nitelenebilir. Diğer taraftan 18 yaşın altında bir öğrencinin bu özelliklere sahip olma durumu, varsayımsal durumlar üzerinden gözlenebilir.

1.4.3. Türetilmiş Ölçme: Ölçülmek istenilen özellik kendisinden farklı iki ya da daha fazla özelliğin arasındaki matematiksel bir bağıntı (dört işlem) yardımıyla belirleniyorsa bu ölçme türüne "türetilmiş ölçme" adı verilir. Örneğin *hız (yol / zaman), yoğunluk (kütle / hacim), zekâ bölümü (zekâ yaşı / takvim yaşı x 100) vb.* özellikler türetilmiş ölçmeye örnek olarak verilebilir.

1.5. ÖLÇÜT TÜRLERİ: Eğitimde ölçüt, mutlak ve bağıl olmak üzere ikiye ayrılır.

1.5.1. Mutlak Ölçüt (Kriter Referanslı Ölçüt): Eğer ölçüt ölçme işleminin yapıldığı grubun özelliklerinden bağımsız olarak belirleniyorsa "mutlak"tır. Mutlak ölçüt genellikle ölçme işlemi öncesinde ilan edilir. Örneğin *dersten geçme notu 50'dir, doktora başvurusu için ALES'ten en az 60 almak gerekir, sürücü ehliyeti sınavında 70 alan başarılı sayılır vb.* Ölçüt daha önceden belirlenmiş ve değişmez standart(lar) olarak saptanmıştır, diğer bir deyişle kuramsal bir yetkinlik standardı esas alınır. Örneğin *"Tam öğrenme kuramı doğrultusunda 100 üzerinden 70, başarılı olma koşuludur."* vb.

Mutlak ölçütün temel alındığı ölçme işlemlerinde bir öğrencinin notu, diğer bir öğrencinin değerlendirmesini (geçme / kalma) etkilemez. Dolayısıyla örneğin başarının ölçüldüğü bir örnekte herkes geçebileceği gibi herkes kalabilir de.

1.5.2. Bağıl Ölçüt (Norm Referanslı Ölçüt): Eğer ölçüt ölçme işleminin yapıldığı grubun belirli bir özelliğine / normuna dayalı olarak belirleniyorsa "bağıl"dır. Bağıl ölçüt ancak ölçme işlemi sonrasında belirlenebilir. Örneğin *aritmetik ortalamaya / ortancaya / moda denk ve üzerinde puan alan geçer, aritmetik ortalama ile yarım standart sapma arasında puan alanlar şartlı geçer, akademik başarısı yüksek ilk on öğrenci proje için seçilecektir vb.* Bağıl ölçütün temel alındığı ölçme işlemlerinde bir öğrencinin başarısı, başka bir öğrencinin değerlendirmesini (geçme / kalma) etkileyebilir. Bu anlamda öğrenciler açısından daha rekabetçi bir eğitsel ortam kaçınılmazdır.

Not 1: Eğer değerlendirme mutlak ölçüte göre yapılıyorsa "mutlak değerlendirme", bağıl ölçüte göre yapılıyorsa "bağıl değerlendirme" adını alır. Örneğin

- *Selim ……. fen lisesine yerleştirilmiştir. → Bağıl değerlendirme*
- *Güzin lisans programları için tercih yapabilecektir. → Mutlak değerlendirme*

Not 2: Eğitsel kararların verilmesinde mutlak ve bağıl ölçütler bir arada da kullanılabilir. Örneğin

- *Yükseköğretim Kurumları Sınavı*
- *KPSS Ortaöğretim Memur Atamaları*
- *Ortalama 40-60 ise mutlak ya da bağıl değerlendirme yapılır; ortalama 60'ın üzerindeyse mutlak değerlendirme, 40'ın altındaysa bağıl değerlendirme zorunludur.*

Not 3: Öğrenci ile ilgili önemli / hayati kararlar verileceği zaman mutlak değerlendirme yapılması gerekir. Örneğin *tek ders sınavları vb.*

Not 4: Başvuran kişi sayısının çok, alınacak kişi sayısının az olduğu (arz-talep dengesizliği olan) sınavlarda yerleştirme / atama vb. bağıl değerlendirme ile yapılmak zorundadır. Örneğin *LGS, YKS, KPSS, TUS vb.*

Not 5: Muafiyet sınavlarında mutlak değerlendirme yapılması gerekir. Örneğin y*abancı dil muafiyet sınavları vb.*



1.6. AMACA GÖRE DEĞERLENDİRME TÜRLERİ: Amaca göre değerlendirme tanıma ve yerleştirme, biçimlendirme ve yetiştirme, değer biçme, rehberlik amaçlı ve program değerlendirme olarak beşe ayrılır.

1.6.1. Tanıma-Yerleştirmeye Yönelik Değerlendirme (Diyagnostik Değerlendirme):

Sınıf içi ölçme ve değerlendirme: Öğretim sürecinin başında, öğrencilerin önceki kritik nitelikteki öğrenmelerini belirlemek amacıyla yapılır. Diğer bir ifadeyle amaç öğrencinin ders bağlamındaki hazırbulunuşluluk düzeyini belirlemektir. Amaç not vermek değildir. Bu amaçla uygulanan ölçme araçlarına da “hazırbulunuşluluk testi” adı verilir.

Okul / kurum geneli ölçme ve değerlendirme: Yine öğretim sürecinin başında, öğrenciyi tanımak ve onu uygun olan programa / kura / sınıfa yerleştirmek amacıyla yapılır. Amaçlanan, not vermek değildir. Hedeflenen, öğrencinin niteliklerine uygun grupların oluşturulmasıdır. Zaman, para ve emek tasarrufu sağlar. Muafiyet sınavları ve seviye tespit sınavları buna örnek gösterilebilir.

1.6.2. Biçimlendirme-Yetiştirmeye Yönelik Değerlendirme (Formatif Değerlendirme): Belirli bir konu / ünite vb. sonunda öğrencilerin öğretime konu olan davranışların ne kadarını kazandığını, diğer bir deyişle ünitedeki öğrenme eksiklerini belirlemek amacıyla yapılır. Amaç yine not vermek değildir. Bu değerlendirme türünde öğrenciye o ünitedeki öğrenme eksikleri bildirilir, bir sonraki kapsama geçmeden öğrenme eksikleri giderilir ya da öğrenme eksiklerinin giderilip giderilmediği denetlenir.

Biçimlendirme-yetiştirme amacıyla yapılacak değerlendirmede kullanılan testlere alanda “izleme testi” ya da “tarama testi” denir. Not verilmeden yapılan kısa sınavlar (quiz), ünite tarama testleri bu amaçla yapılan ölçme ve değerlendirmeye örnek olarak verilebilir.

Bir ünitedeki öğrenmeler bir başka ünite için ön öğrenme hâline gelebilmektedir. Ön öğrenme eksiklerinin giderek artması öğrencinin öğrenemez hâle gelmesine kadar gidebilir. Blooma'a göre bir ünitedeki öğrenme eksikleri belli bir düzeyin üzerindeyse o öğrenme açığı kar topu gibi devam eder. Öğrencinin öğrenemez hâle gelmemesi için ünite sonlarında kontrol mekanizmalaları çalıştırılmalıdır.

1.6.3. Değer Biçmeye / Düzey Belirlemeye Yönelik Değerlendirme (Summatif Değerlendirme): Belirli bir öğretim sürecinin sonunda, not vermek[8] amacıyla yapılan değerlendirmedir. Değer biçmeye yönelik değerlendirme, öğrencinin belirli bir öğretim süreci sonundaki durum tespitidir çünkü artık hedeflenen öğretim sürecinin sonuna gelinmiştir. Bu amaçla uygulanan testlere “alanda erişi testi” denir. Bitirme sınavları, sertifika sınavları, üniversitedeki vize ve finaller, eğitim-öğretim kurumlarında not vermek amacıyla uygulanan yazılı yoklamalar bu ölçme ve değerlendirme türüne örnek olarak verilebilir.

İlk üç değerlendirme türü öğrenci başarısına odaklıdır, diğer bir ifadeyle öğrencinin akademik başarısının ölçülüp değerlendirilmesi söz konusudur. Ancak eğitimde yalnızca akademik başarının ölçülüp değerlendirilmesi ile ilgilenilmez. Öğrencilerin yetenekleri, ilgileri, tutumları, kaygıları, öğrenme kapasiteleri ve karşılaşılan güçlükler ile uygulanan eğitim programının etkililiğinin değerlendirilmesi de çok önemlidir.

1.6.4. Rehberlik Amaçlı Değerlendirme: Rehberlik amaçlı değerlendirme kabaca iki bağlamda ele alınabilir. Bunlar özel eğitim ve mesleki rehberlik hizmetleridir.

[8] Öğrenci yapılan her ölçme işleminden sayı değeri ya da sembol alır. Her sayı değeri not anlamına gelmez. Eğer puan öğrencinin geçme/kalma vb. bir kararla kullanılıyorsa o ölçüme “not” adı verilir.



Özel eğitim: Belirli alanlarda özel gereksinimi olan çocuklara uygun eğitim ortamları ve/veya programı uygulamak amacıyla ölçme ve değerlendirme süreçleri işe koşulmaktadır. Özel gereksinim kavramı zihinsel gerilikten öğrenme güçlüğüne, otizmden diğer spesifik gerilik alanlarına, üstün zekâdan işitme/görme vb. engel alanlarına kadar çok geniş bir yelpazede ele alınır. Özel gereksinimi olan çocuklara uygun eğitim olanaklarının sunulabilmesi için çocuğun ihtiyacı olan alanlar, eğitsel tanılama amacıyla genellikle standart testler veya ölçüt bağımlı testler (beceri öğretimde) ile belirlenmeye çalışılır. Bu tür testlerin sonuçları doğrultusunda, tıbbi ve eğitsel / gelişimsel olarak tanı alan çocuklar için bireysel eğitim programları (BEP) hazırlanır. Eğitsel değerlendirme bağlamında RAM'lar bu konudaki tek yetkilidir.

Mesleki rehberlik: Öğrencilerin alan seçmelerinde ve özellikle bir üst öğretim kurumuna yönlendirilmesinde, özelliklerine uygun alan / meslek seçmeleri hem öğrenci hem de ülkenin insan gücü kaynağının doğru planlanması açısından çok önemlidir. Bu amaçla öğrencilerin kendi yetenekleri, kişilik özellikleri, ilgileri vb. ile özelliklerini belirlemek amacıyla standart testlerden yararlanılmaktadır. Böylece öğrencinin kendi özellikleri hakkındaki farkındalıklarının artırılması ve dolayısıyla buna uygun alan / meslek seçmeleri sağlanmış olur.

1.6.5. Program Değerlendirme: Eğitim programları bileşenlerinin yapısı ve niteliğine ilişkin olarak sistematik bilgi toplama ve değerlendirme sürecidir. Program değerlendirme eğitim programlarının planlanması, mevcut programlar ve/veya ürünlerin etkililiğinin değerlendirilmesi ve eğitim programlarının ve/veya ürünlerin geliştirilmesi amacıyla yapılır. Eğitim programlarını değerlendirme, değerlendirilmekte olan programın boyutlarıyla ilgili sorularla yönlendirilen uygulamalı bir etkinliktir ve bilimsel araştırma sürecinin adımlarıyla büyük ölçüde benzerlik gösterir. Program değerlendirmede program taslağının hazırlanması, programın uygulanması ve programın etkisine ilişkin yargılar elde edilir.

2. BİR ÖLÇME ARACINDA BULUNMASI GEREKEN PSİKOMETRİK NİTELİKLER

Doç. Dr. Güçlü ŞEKERCİOĞLU

Ölçme her zaman belirli bir amaç için yapılır. Amaç ölçmeye konu olan özellik bakımından kişiler hakkında değerlendirme yapmak ve elde edilen değerlendirme sonuçlarına göre belirli kararlar vermek için yapılır. Verilen kararların doğru ve isabetli olması, değerlendirmenin dayandığı ölçümün özellikle doğrudan ilgili ve olabildiğince az hatalı olmasına bağlıdır. Psikolojik özelliklerin ölçülmesinde genellikle ölçme araçları kullanılır ve bu araçlardan elde edilen puanların hatasız ya da az hatalı olması ve bu araçların belirli psikometrik niteliklere sahip olması gerekir. Bunlar:

Geçerlilik ölçmek istenilen özelliğin, başka özelliklerle karıştırılmadan, doğru ve tam olarak ölçülebilmesidir. Diğer bir ifadeyle ölçme aracından elde edilen puanın amaca hizmet etme derecesidir.

Güvenilirlik ölçme işleminden elde edilen puanların tesadüfi hatalardan arınık olma derecesidir. Diğer bir ifadeyle puanların kararlı, tutarlı ve duyarlı olmasıdır.

Kullanışlılık ise ölçme aracının geliştirilmesinin, uygulanmasının ve puanlanmasının kolay ve ekonomik (zaman, para, emek, araç gereç vb. açıdan) olması ile ilgilidir.

Bir ölçme aracından elde edilen puanların hatasız ya da az hatalı olması geçerli ve güvenilir ölçme yapmanın tek koşuludur. Ölçme işlemlerinde hataları en aza indirebilmek için öncelikle hataların tanımlanması gerekir ki ölçme işlemini yapan kişi neye müdahale edeceğini, hangi durumlara karşı önlem alması gerektiğini bilsin.

2.1. HATA: Yalnızca eğitimde değil bütün bilim dallarında, hatta günlük yaşamımızda dahi yapılan ölçme işlemlerinde bile ölçme sonuçlarına hata karışma olasılığı vardır. Ölçme yapan kişilerin dikkati ve titizliği zamandan zamana değişebilir, ölçme aracına ilişkin bazı sorunlar söz konusu olabilir, ölçülen özelliğin doğası gereği bazı sıkıntılar olabilir, test katılımcısı yeterince güdülenmemiş olabilir, ölçme işleminin yapıldığı ortamdan kaynaklı sorunlar vb. olabilir.



Ölçmede gözlenen bir özelliğin gerçek değeri (sayı, sembol ya da sıfat) bulunmak istenir. Ancak ölçmeye karışan çeşitli hatalar nedeniyle gerçek değer (puan) ölçme yoluyla doğrudan elde edilemez. Psikometride gerçek puan kuramı olarak ele alınan bu kavram basit bir eşitlik ile gösterilir[9].

Bu çerçevede hatalar; kaynağı, yönü ve miktarı göz önünde tutularak sabit, sistematik ve tesadüfi olmak üzere üç türde incelenebilir:

2.1.1. Sabit Hata: Miktarı ölçmeden ölçmeye değişmeyen, diğer bir deyişle her ölçme işlemine aynı miktarda karışan hatalardır. Örneğin marketteki terazi, üzerinde herhangi bir nesne yokken terazi -120 g gösteriyorsa ne tartılırsa tartılsın 120 g eksik ölçülecektir. Bir öğretmen, sınavında herkese 10 puan fazla veriyorsa yine karışan hata sabit olacaktır.

2.1.2. Sistematik Hata: Ölçülen büyüklüğe, öğretmene ya da ölçme koşullarına göre miktarı değişen hatalardır. Örneğin marketteki terazi, her bir kilogramda 120 g eksik tartıyorsa üzerine konulan nesnenin ağırlığı arttıkça hata miktarı da artacaktır. Bir öğretmen, sınavında yazısı kötü olandan puan kırıyorsa yine karışan hata sistematik olacaktır.

Gerek sistematik gerekse sabit hatalarda ölçme sonuçlarına karışan hata miktarı, yönü ve kaynağı bellidir. Bu nedenle bu tür hataların ölçme sonuçlarına karışmasını engellemek ya da bu tür hataları düzeltmek görece daha kolaydır.

2.1.3. Tesadüfi (Rastlantısal) Hata: Şansla ortaya çıkan ne yönde ve ne ölçüde karıştığı genellikle bilinemeyen hatalardır. Ölçmelere tek yönlü olarak karışmaz, ölçme sonuçlarına bazen pozitif bazense negatif yönde etki eder. Sabit ve tesadüfi olmayan hatalar, tesadüfi değişken olma özelliğine sahiptir. Bu nedenle psikometride hata kuramı tesadüfi hatalar üstüne kurulmuştur. Tesadüfi hataların genellikle dört kaynağı olduğu kabul edilir. Bunlar:

2.1.3.1. *Ölçme işlemini yapan kişiden / öğretmenden kaynaklanan hata*: Öğretmenin test etme sürecine ilişkin davranışlarını iki aşamada ele alabiliriz:

<u>Ölçme işlemi sürecinde</u>: Test katılımcısının dikkatini dağıtacak ve/veya kaygısını artıracak davranışlar vb.

<u>Ölçme işlemi sonrasında</u>: Puanlamadaki dikkat ve titizliğin zamandan zamana değişmesi, yorgunluk, öncelik-sonralık yanılgısı, maddi hata vb.

2.1.3.2. *Ölçme aracından kaynaklanan hata*: Ölçme araçları hazırlanırken maddelerin iyi ifade edilmemesi, test katılımcılarının yanlış anlamalarına ve dolayısıyla hataya neden olacaktır. O nedenle hangi özellik ölçülürse ölçülsün, araçta kullanılan dilin dil bilgisi, imla kurallarına uygun olması; maddelerde anlatım bozukluğu bulunmaması, muğlâk ya da belirsiz (müphem) ifadelerin bulunmaması gerekir.

Ölçme aracından kaynaklanabilecek bir başka hata kaynağı ise aracın yapısı ile ilgilidir. Ölçtüğü özellik ve kapsam bakımından homojen (benzeşik) maddelerden oluşan bir araç, heterojen (ayrışık) maddelerden oluşan bir testten daha güvenilirdir. Diğer taraftan ölçülecek özellikler evrenini yeterince temsil etmeyen maddelerden oluşan bir araçtan elde edilen puanlara da hata karışmaktadır.

Ölçme aracının yapısıyla ilgili diğer bir konu ise aracın uzunluğudur. Madde sayısı ile güvenilirlik arasında doğru orantılı bir ilişki vardır ancak bu sonsuz bir doğru orantı anlamına gelmez. Ölçme aracındaki madde sayısı arttıkça bireyde yorgunluk, dikkat azalması vb. etkenler nedeniyle hata miktarı artacaktır.

Ölçme aracıyla ilgili diğer bir faktör ise bilişsel özellikleri ölçen testlerde eğer seçenek verilmişse şans başarısı[10] karışma olasılığıdır. Bu durum ölçülen özelliğin gerçekte olduğundan yüksek görünmesine yol açmaktadır.

---

[9] Bu kavram ve eşitlik güvenilirlik başlığı altında ele alınmıştır.

[10] Salt tahminle doğru yanıtı bulma olasılığı



2.1.3.3. *Bireyden / öğrenciden kaynaklanan hata*: Bireylerin ölçme işlemi sürecinde içinde bulundukları fiziksel, fizyolojik ve psikolojik durumlar test puanına etki eder. Uykusuzluk, açlık, hastalık, ağrı, motivasyon eksikliği, kaygı vb. faktörler nedeniyle bireyler gerçek puanlarından uzaklaşabilirler. Ayrıca grubun homojen ya da heterojen olması da hata kaynağını arttıran ya da azaltan bir etkiye sahiptir. Daha heterojen gruplarda güvenirlik katsayısı artarken daha homojen gruplarda güvenirlik katsayısı azalır.

2.1.3.4. *Fiziksel ortamdan kaynaklanan hata*: Ölçme işleminin gerçekleştiği fiziksel ortama ilişkin bazı etmenler de ölçme sonuçlarına hata karıştırabilir. Sıcaklık, ışık, ses, koku, görsel uyarıcıların fazlalığı, ergonomi vb. etmenler hata miktarını artırabilir. Bu anlamda testler bireylere bu etmenler açısından eşit ve standart koşullar altında uygulanmalıdır, aksi takdirde hata miktarının artması kaçınılmazdır.

2.2. GEÇERLİLİK, GÜVENİLİRLİK VE HATA İLİŞKİSİ: Geçerlilik tüm hata kaynaklarından etkilenirken klasik test kuramına göre güvenilirlik yalnızca tesadüfi hatalardan etkilenir.



Şekil 1: Geçerlilik, güvenilirlik ve hata ilişkisi

Bu nedenle güvenilirlik, geçerlilik için bir ön şart ancak yeterli şart değildir. Diğer bir ifade ile bir testin güvenilir olması onun geçerli olacağı anlamına gelmez ancak bir test geçerli ise büyük olasılıkla güvenilirdir.

Bir ölçme aracının geçerliliği ve güvenilirliği diye bir şey yoktur, ölçme araçlarından elde edilen puanların geçerliliği ve güvenilirliği diye bir şey vardır. Ölçme araçlarının psikometrik nitelikleri şu durumlara göre değişebilir: a. Ölçme amacının değişmesi, b. Uygulama grubunun değişmesi, c. Dilin eskimesi / değişmesi, d. Maddelerde ve/veya alt ölçeklerde yapılan değişiklikler, e. Farklı kültürler, f. Kuramsal bilgi birikiminde değişiklikler.

Geçerlilik ve güvenilirlik bir varlık-yokluk sorunu değil, derece sorunudur. Bu nedenle bir testten elde edilen puanlar yüksek düzeyde geçerli, orta düzeyde geçerli, düşük düzeyde geçerli; yüksek düzeyde güvenilir, orta düzeyde güvenilir, düşük düzeyde güvenilir biçiminde nitelendirilir.

Bir ölçme aracının puanlarından yapılan değerlendirmenin doğruluğu araştırmayı gerektirir. Bu puan gerçekten ölçmek istediğimiz özelliği gösteriyor mu? Aracın ölçmek istediğimizi ölçüp ölçmediği bir araştırma işidir. Araştırma ile değerlendirmenin doğruluğu incelenir. Bir testin şöhretli olması onun geçerli olduğu anlamına gelmez. Modern test kuramına göre geliştirilen ölçekler eleştiriye açıktır. Böylece testlerin niteliği ve test bilgisi gelişir.

2.3. KORELASYON: Geçerlilik ve güvenilirliği belirlemeye yönelik yöntemler genellikle korelasyon temelli analizlere dayanır. Bu nedenle bu aşamada kısaca korelasyon kavramına değinmek gerekmektedir. Korelasyon (co-relation), en az iki değişken arasında karşılıklı bir ilişki bulunup bulunmadığı, eğer ilişki varsa bu ilişkinin yönü ve miktarı hakkında bilgi veren istatistik bir tekniktir. “r” ile sembolize edilir. Örneğin

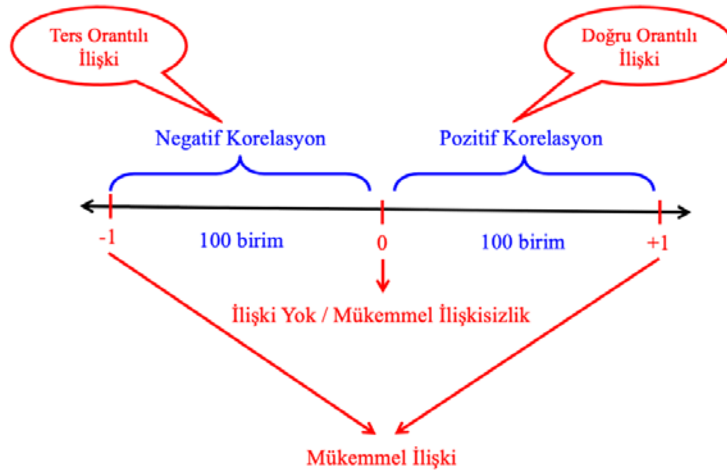
- *Öğrencilerin derse ilişkin tutumları ile ders başarıları arasında bir ilişki var mıdır?*



- *Öğretmenlerin ders saati yükleri ile iş doyumları arasında bir ilişki var mıdır?*
- *Saç uzunluğu ile zekâ arasında bir ilişki var mıdır?*

Korelasyon -1 ile 1 arasında değer alır. Bu iki değer arasında matematiksel olarak sonsuz birim vardır ancak kullanışlılık açısından 100 birim negatif korelasyonda, 100 birim pozitif korelasyonda, sıfır ile birlikte toplam 201 birimlik bir skala üzerinden değişkenler arasındaki ilişkinin yönü ve miktarı değerlendirilir.

Pozitif korelasyon iki değişken arasında doğru orantılı ilişki anlamına gelir. Örneğin ders çalışma süresi ile sınav notu ya da gelir ile tüketim arasında pozitif bir korelasyonun elde edilmesi beklenir. Negatif korelasyon, iki değişken arasında ters orantılı ilişki anlamına gelir. Örneğin ders süresi ile dikkat ya da yükseklik ile sıcaklık arasında negatif bir korelasyonun elde edilmesi beklenir. ∓1 mükemmel korelasyon anlamına gelir ancak 1 mükemmel doğru orantılı ilişki anlamına gelirken, -1 ise mükemmel ters orantılı ilişki anlamına gelir. Diğer taraftan sıfır korelasyon iki değişken arasında sistematik bir ilişkinin olmadığı anlamına gelir. Bu duruma mükemmel ilişkisizlik de denir. Örneğin öğretmenlerin boyları ile aylık ücretleri arasındaki korelasyonun sıfır olması beklenen bir durumdur.

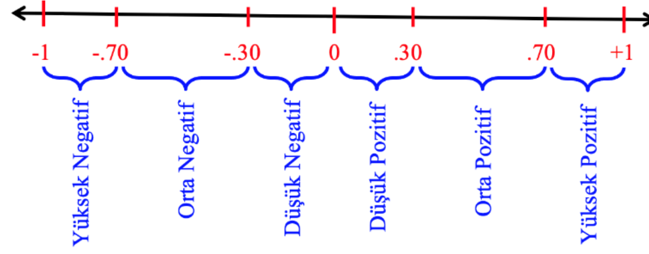


Şekil 2: Korelasyonun temel kavramları

Korelasyon katsayısı hakkında kabaca iki tür belirleme yapmak gerekir. Bunlardan birincisi yön, diğeri ise miktardır. Yön, negatif ya da pozitif olarak değerlendirilirken miktar, kabaca düşük, orta ya da yüksek olarak nitelendirilir. Miktar için kesin sınırlar olmamakla birlikte alanyazında genellikle kabul gören aralıklar Şekil 3'te gösterilmiştir.







Şekil 3: Korelasyonun miktarı

Not 1: Korelasyon mutlak değer olarak değerlendirilmelidir. Bir korelasyon katsayısının negatif ya da pozitif olması büyüklük-küçüklük belirtmez, yön bildirir.

Not 2: Korelasyon katsayısı ile neden-sonuç ilişkisi kurulamaz. Değişkenler arasında doğru ya da ters orantılı bir ilişki olması, söz konusu değişkenler arasında bir neden-sonuç ilişkisinin varlığı anlamına gelmez.

2.4. GEÇERLİLİK SORGULAMA YÖNTEMLERİ: Geçerlilik sorgulama yöntemleri aşağıda Şekil 4'te gösterilmiştir.



Şekil 4: Geçerlilik yöntemleri sınıflaması

2.4.1. KAPSAM GEÇERLİLİĞİ: Kapsam geçerliliği özellikle başarı testlerinde aranan bir geçerlilik sorgulamasıdır. Eğitimde öğretmen bir program dâhilinde önceden belirlenen davranışları kazandırmayı amaçlar. Belirli bir zaman sonra davranışların ne kadarının kazanıldığını belirlemek isteyen öğretmen, geliştirdiği testte öğretime konu olan içeriğin testte ne ölçüde temsil edildiğini belirlemek durumundadır. Bu anlamda kapsam geçerliliği bir testin ölçülmek istenen davranışları ne derece kapsadığıyla ilgilidir. Ne amaçla kullanılırsa kullanılsın test, kapsamı açısından ölçmeye konu olan davranışları yeterli ve dengeli bir biçimde temsil etmelidir. Bir testin kapsam geçerliliğinin yüksek olduğunun söylenebilmesi için

1. Testteki soruların / maddelerin ölçülecek özellikler evrenini (konu kapsamını / içeriği) yeterli ve dengeli bir biçimde ölçüyor olması ve
2. Her bir sorunun / maddenin ölçmek istediği özelliği doğrudan ölçmesi, diğer bir deyişle kazanımla doğrudan ilgili olması gerekir.

Testin kapsam geçerliliğinin yüksek olduğunun söylenebilmesi için bu iki kriter birden sağlanmalıdır.

Kapsam geçerliliği sorgulama yöntemleri:

2.4.1.1. Mantıksal / rasyonel yöntemler: Bu yöntemler belirtke tablosu hazırlanması ve uzman görüşüne başvurulmasıdır.



Belirtke tablosunun hazırlanması: Bir kapsam geçerliliği sorgulamasında öncelikle ölçmeye konu olan kapsam dâhilinde davranışların belirlenmesi gerekir. Bu noktada en çok kullanılan yöntemlerden biri belirtke tablosu hazırlamaktır. Öğretmen satırda davranışları, sütunda hedefleri yazar; belirlediği madde sayısı doğrultusunda davranış ve hedefleri yeterli ve dengeli bir biçimde temsil eden alanları seçer.

Tablo 3. *Belirtke tablosu örneği*

Hedefler / **Konular**	**Bilgi**	**Kavrama**	**Uygulama**	**Analiz**	**Sentez**	**Değerlendirme**	**Toplam**
1. Ölçme ve değerlendirmenin temel kavramları							**%43.3**
1.a. Ölçme ve değerlendirmenin eğitim sistemi içindeki yeri		1				1	**2**
1.b. Eğitimde paradigma değişimi ve bu değişimin ölçme ve değerlendirme uygulamalarına yansımaları	1			2			**3**
1.c. Ölçme, değerlendirme ve ölçütün kavramsal temelleri		2					**2**
1.d. Ölçme türleri		1					**1**
1.e. Ölçüt türleri	1					1	**2**
1.f. Değerlendirmenin amaca göre türleri		1			1	1	**3**
2. Bir ölçme aracında bulunması gereken psikometrik nitelikler							**%56.7**
2.a. Geçerlilik, güvenilirlik ve kullanışlılığın kavramsal temelleri		1					**1**
2.b. Ölçmede hata kavramı	1						**1**
2.c. Hata türleri		1		1			**2**
2.d. Geçerlilik ve güvenilirlik ile hata ilişkisi						1	**1**
2.e. Geçerlilik ve güvenilirlikle ilgili temel kavramlar		1	2				**3**
2.f. Geçerlilik sorgulama yöntemleri		2		1			**3**
2.g. Güvenilirlik kestirim yöntemleri	1	1	1			1	**4**
2.h. Geçerliliği ve güvenilirliği artırma yolları		1		1			**2**
TOPLAM	**4**	**12**	**4**	**5**	**1**	**5**	**30**

Uzman görüşüne başvurulması: Uygulamada uzman ile kastedilen öncelikle ölçme ve değerlendirme tekniklerini de bilen bir alan uzmanıdır. Uzmana belirtke tablosu ve maddeler sunulur, uzmandan soruların / maddelerin konu kapsamını yeterli ve dengeli bir biçimde ölçme durumunu ve soruların / maddelerin kazanımlarla doğrudan ilgili olma durumunu değerlendirmesi istenir.

2.4.1.2. İstatistiksel yöntemler: Uzmanlardan alınan dönütler betimsel / muhakemeye dayalı bir yolla çözümlenebilir ya da uzmanların *"uygundur / uygun değildir"* ya da *"uygundur / düzeltme gerekir / soru kullanılmamalıdır"* vb. biçimde değerlendirme yapması istenebilir. Eğer uzmandan ikinci yolla dönüt istenmişse uzmanlar arasında uyuşum olup olmadığı, çeşitli istatistiksel yöntemlerle test edilir. Alanda uzmanlar arasında uyumu test eden pek çok indeks bulunmaktadır.

2.4.2. ÖLÇÜT DAYANAKLI GEÇERLİLİK: Ölçme aracından elde edilen puanların ölçüt bir puanla (testin tahmin etmeye çalıştığı ve geçerliliği yüksek bir puan) karşılaştırılarak geliştirilen ölçme aracının geçerliliğine ilişkin nitelendirme yapılır.

2.4.2.1. YORDAMA GEÇERLİLİĞİ: Yordama, tahmin demektir ancak her tahmin yordama değildir. Bir tahminin yordama olabilmesi için elde geçerli ve güvenilir bir veri olması ve bu verinin sınanabilir, sayısal nitelikte, belirli analizlere tabi tutuluyor olması gerekiyor. Diğer bir deyişle yordama, eldeki bu nitelikteki veriden yola çıkarak geleceğe, henüz gerçekleşmemiş bir olguya ilişkin yapılan tahmindir.

Ölçme araçlarının çoğunda bireylerin gelecekteki davranışlarının kestirilmesi söz konusudur. Yordama geçerliliği, ölçme aracının bu kestirim işini ne ölçüde doğru ve isabetli



yaptığı ile ilgili bir sorgulamadır. Özellikle iki amaçla uygulanan ölçme araçlarının yordama geçerliliğinin yüksek olması istenir. Bunlar seçme ya da yönlendirme amaçlı kullanılan testlerdir. Seçme amaçlı testlere YKS, KPSS, LGS vb. araçlar; diğer taraftan yönlendirme amaçlı testlere ise alan seçiminde ve/veya bir üst öğretim kurumuna yönlendirmek için uygulanan yetenek testleri, ilgi envanterleri, mesleki kişilik envanterleri vb. örnek olarak gösterilebilir.

Yordama geçerliliğinde ölçme aracından elde edilen puanlar, ölçme aracının tahmin ettiği puanla (ölçüt puan) karşılaştırılır ve tahminin ne ölçüde doğru olduğu belirlenmeye çalışılır. Ölçüt puan testin tahmin etmeye çalıştığı özelliktir. Yordama geçerliliğinde ölçüt puan gelecekte belli olacağından beklemek gerekmektedir. Yordama geçerliliğinde geçerlilik katsayısının 0 ile 1 arasında değişmesi beklenir.

	YKS Puanı	**Akademik Ortalama**
Tarık	438.636	3.54
Esra	348.545	2.45
Begüm	456.344	3.39
Ozan	399.987	2.23
Defne	382.654	2.34
……	480.334	3.45
……	287.987	1.67
……	381.954	3.88
……	250.897	3.91
Umut	343.019	3.26

Ölçüt gelecekte belli olacağından beklemek gerekmektedir.
Korelasyon (r) 1'e yakın
Geçerlilik Katsayısı

Şekil 5: Bir yordama geçerliliği çalışma örneği

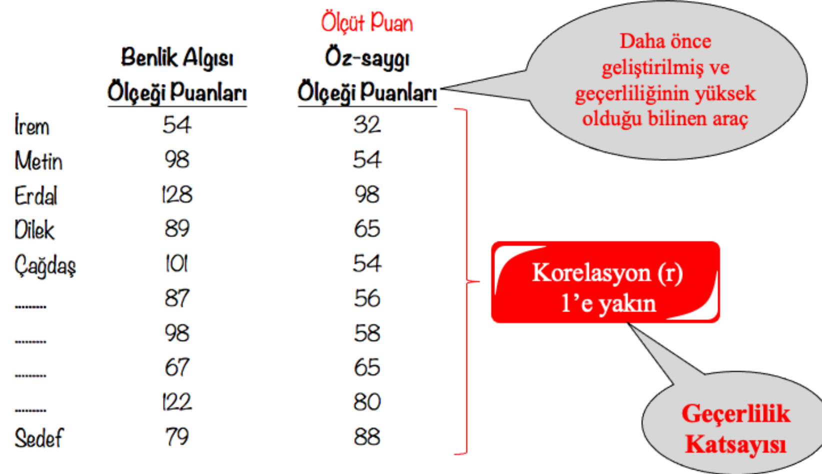
Yordama geçerliliğinde en zor ve önemli nokta ölçütün doğru bir biçimde belirlenmesidir. Ölçütün belirlenmesinde şu noktalar dikkate alınmalıdır: Ölçüt(ün);

1. Ölçme aracının yordamaya çalıştığı değişkenle doğrudan ilgili olmalı, ölçme aracı hangi özelliği kestirmeyi amaçlıyorsa onun doğrudan bir temsili olmalıdır.
2. Kararlı olmalı, günden güne değişmemelidir. Açıktır ki kendisi kararsız olan bir özellik hiçbir araçla yordanamaz.
3. Bireylerin özelliğini gerçekten yansıtan nesnel ve güvenilir bir ölçüt olmalıdır. Söz gelimi okulda alınan notlar bir ölçüt olarak alınmışsa öğrencilere verilen notlara başarının dışındaki etmenler etki etmemelidir.
4. Elde edilmesi kolay ve ekonomik olmalıdır.

2.4.2.2. ZAMANDAŞ GEÇERLİLİK: Bu geçerlilik türü alanyazında hâlihazır geçerlilik, benzer ölçekler geçerliliği, uygunluk geçerliliği adı ile de anılmaktadır.

Zamandaş geçerlilik sorgulamalarında ölçüt puan eş zamanlı olarak elde edilebilir. Geliştirilen ölçme aracı ile ilişkili olabilecek nitelikleri ölçen ve geçerliliği yüksek bir aracın puanı ölçüt puan olarak ele alınabilir. Daha sonra geliştirilen ölçme aracı ve ölçüt araç aynı zamanda uygulanıp iki araçtan elde edilen puanlar arasındaki ilişki incelenir.





Şekil 6: Bir zamandaş geçerlilik çalışma örneği

Zamandaş geçerliliğinde geçerlilik katsayısının –1 ile 1 arasında değişmesi beklenir. Ölçüte bağlı olarak geçerlilik katsayısı ∓1'e yaklaştıkça artar, 0'a yaklaştıkça düşer.



Şekil 7: Geçerlilik katsayısı

Zamandaş geçerlilikte de en zor ve önemli nokta ölçütün doğru bir biçimde belirlenmesidir. Ölçütün belirlenmesinde şu noktalar dikkate alınmalıdır: Ölçüt(ün);

1. Ölçme aracının ölçmeye yöneldiği özellikle doğrudan ilişkili olmalıdır. Bu ilişki doğru orantılı ya da ters orantılı olabilir.
2. Geçerliliği yüksek olmalıdır. Geçerliliği yüksek olmayan bir ölçüt puanla bakılacak korelasyonun düşük olması kaçınılmazdır.

2.4.3. YAPI GEÇERLİLİĞİ: Beşerî bilimlerde atılganlık, güvensizlik, içe dönüklük vb. özelliklerin ölçülmesi amacıyla ölçme aracı geliştirmek için önce, söz konusu olan özellik ya da özellikleri belirlemek yani yapıyı tanımlamak; daha sonra, tanımlanmış yapıdan sınanabilir denenceler çıkarmak ve çıkarılan denenceleri sınamak için deneysel ve istatistiksel çalışmalar yapmak gerekmektedir. Yapı, birbirleriyle ilgili olduğu düşünülen belli ögelerin ya da ögeler arasındaki ilişkilerin oluşturduğu bir örüntüdür. Bu anlamda, bir testin yapısını geçerleme süreci, temelde testin maddelerine verilen yanıtlar arasındaki ilişkilerin analizine dayanır.

Yapı geçerliliği, bir testin dayandığı kuramsal temelleri ne derece iyi örneklediğiyle ilgilidir. İnsan davranışları ve özellikleri çoğunlukla soyut bir yapıya sahiptir. Zekâ, sevgi, merak, sosyal uyum ve ruh sağlığı gibi soyut kavramları psikolojik testlerle ölçebilmek için önce bu kavramların içeriğinin bilinmesi gerekir. Birey ne tür davranışlar gösterdiği zaman hangi niteliklerin varlığına ya da yokluğuna karar verilebileceğine ilişkin ölçütler geliştirilmesi gerekmektedir. Yapı geçerliliği bir yandan testin ölçtüğü niteliklerin neler olduğunu araştırma, diğer yandan testi alan kişilerin elde ettikleri puanların ne anlama geldiğini açıklama çabalarıyla ilgilidir. Örneğin bir kişi, geleneksel aile biçimi ile çocuk yetiştirme biçimi arasındaki ilişkiyi incelemek amacıyla bir ölçek geliştirip bu ölçeğin yapı geçerliliğini ortaya koymak istediğinde,



geleneksel aile yapısı ve çocuk yetiştirme kavramlarının ne anlamlara geldiğini, ölçme aracındaki maddelerin bu anlamlara uygunluğunu araştırarak yapı geçerliliği konusunda karar verebilir.

2.5. GÜVENİLİRLİK:

Güvenilir bir ölçme aracı, aynı özellikle ilgili olarak arka arkaya yapılan ölçmelerde yaklaşık olarak aynı sayısal sonucu verir; diğer bir ifadeyle bir test, aynı gruba iki ya da üç kez uygulandığında gruptaki her bir kişi bütün uygulamalarda yaklaşık olarak aynı puanı almalıdır. Bir testin ölçmek istediği özelliği ölçebilmesi için o testin söz konusu olan özelliği kararlı olarak ölçmesi gerekir. Güvenirlik çalışmalarının odak noktası şudur: *“Eğer kişi iki defa teste tabi tutulursa iki testten aldığı puanlar birbirine benzer midir ya da birbirine ne kadar yakındır?”* Bu noktada gözlenen puan, gerçek puan, ölçmenin standart hatası ve güvenirlik katsayısı kavramlarına değinmek gerekmektedir.

Örneğin *Murat, üç dakikalık bir sözcük çalışmasında 162 sözcük ya da diğer bir ifadeyle dakikada 54 sözcük yazmıştır. Bu puan Murat’ın becerisini ne ölçüde yansıtmaktadır. Varsayalım ki dakikada 50 sözcük yazılması yeterli kabul edilmektedir. Murat bu düzeyin gerçekten üzerinde midir? Murat geçen hafta dakikada 45 sözcük yazmıştır. Murat’ın bugünkü puanı olanı olan 54, onun kendisini geliştirdiğinin göstergesi midir ya da bir değişim dalgalanması mıdır?*

İki ölçüm arasındaki uyumsuzluğun birçok nedeni olabilir. Bir hareketten diğerine, *“dikkat ve çaba”* değişebilir. Özellikle uzun periyotlarda puan değişmesi, fiziksel büyüme, öğrenme ya da sağlık ve kişilikteki değişimlerden kaynaklanabilir. Yine ikinci ölçümde daha açık soruların kullanılması diğer bir faktör olabilir. İki ölçüm arasındaki puan farklarını yorumlayabilmek için gerçek puan kuramına bakalım.

Psikometride hata kavramı istenmeyen değişkene işaret eder. Ölçme hataları giderilene kadar sürdürülmeli ve böylece gerçek puan elde edilmelidir. Ancak davranış örneği sınırlı olduğu için gözlenen puan gerçek puandan farklılık gösterir. Buradaki farklılık ölçme hatasıdır. Geleneksel olarak hataların varlığı gözlenen puanın gerçek puandan yüksek ya da düşük olmasına neden olur.

Örneğin *yarışlara hazırlanan bir koşucu, bir mesafeyi farklı zamanlarda 23.7, 24.0, 24.2,... 25.1, 25.2 saniyelerde koşmuş olsun. Bu ölçümlerin ortalaması 24.7 ise gerçek puanı 24.7’dir. Koşucu bu puana daha önce 23.7 saniyede koştuğu ölçümü göstererek itiraz edebilir. Bu durumda koşucuya 23.7 ve 25.2 saniyede koştuğu durumlara birçok faktörün etkisinin olabileceği ve 23.7 saniyede tekrar koşmasının belki de hiç olanaklı olamayacağını, gerçek puanının 24.7 olduğunu söyleyebiliriz.* Bu söylem **gerçek puan kuramı**na dayanır.

$$\mathbf{X = T + E}$$

X = Bireylerin ölçme aracından elde ettiği gözlenen puanı

T = Bireylerin gözlenemeyen gerçek puanı

E = Ölçmeye karışan hata miktarı

Varsayımsal olarak bir ölçme işleminde hata miktarı sıfır ise eşitlik X = T + 0 ve dolayısıyla X = T olmuş olur. Diğer bir ifade ile gözlenen puan, gerçek puana eşit olmuş olur. Ölçme işleminde hata miktarı arttıkça gözlenen puanın gerçek puana olan farkı artmaya başlar. Ölçmenin standart hatası

$$S_h = s\sqrt{1 - r_x}$$



formülü ile bulunur.

S_h= *Ölçmenin standart hatası*

s= Standart sapma

r_x= *Güvenilirlik katsayısı*

Bir ölçme işleminde standart hatanın düşük olması, formülden de anlaşılacağı üzere, güvenilirlik katsayısının yüksek, standart sapmanın görece düşük olmasına bağlıdır. Örnek: Bir ölçme işleminde s=8 ve r_x=.75 ise $S_h=8\sqrt{1-.75}=4$. Bu sonuç bireylerin puanlarına –/+ 4 puan hata karıştığı anlamına mı gelir? Gerçek puanın tahmin edilmesinde genellikle üç olasılık değeri kullanılır. Bunlar yaklaşık %68, %95 ve %99'dur. %68 olasılık için bireyin puanından bir standart hata çıkarılır ve puanına eklenirken %95 olasılık için bireyin puanından iki standart hata çıkarılır ve puanına eklenir, son olarak %99 olasılık için bireyin puanından üç standart hata çıkarılır ve puanına eklenir.

%68 olasılık için $X \mp 1 \times S_h$

%95 olasılık için $X \mp 2 \times S_h$

%99 olasılık için $X \mp 3 \times S_h$

Dolayısıyla bir kişi 100 üstünden 50 puan almış ve S_h=4 ise

%68 olasılık: $X \mp 1 \times S_h \Rightarrow 50 \mp 1 \times 4$=Bireyin gerçek puanı %68 olasılıkla 46-54 arasında değişir.

%95 olasılık için $X \mp 2 \times S_h \Rightarrow 50 \mp 2 \times 4$=Bireyin gerçek puanı %95 olasılıkla 42-58 arasında değişir.

%99 olasılık için $X \mp 3 \times S_h \Rightarrow 50 \mp 3 \times 4$=Bireyin gerçek puanı %99 olasılıkla 38-62 arasında değişir.

Ölçmenin standart hatası yükseldikçe ölçme işlemi bireylere rastgele puan vermekten farklı olmayacaktır. Bu nedenle ölçme işlemlerinde güvenilirliğin kestirilmesi çok önemlidir.

Güvenilirlik katsayısı 0 ile 1 arasında değişir. Güvenilirlik katsayısı için alanyazında genellikle 0.70 ve üstü ölçütü kabul edilir. Ancak bu değerin de çok yüksek bir değer olmadığı, 0.70-0.80 arası güvenilirliğin ancak ön bilgi elde etmek amacıyla kullanılabileceği ifade edilir. Genel yetenek gibi bilişsel özellikleri ölçen testlerde güvenilirlik katsayısının 0.90 ve üzerinde olması arzu edilir.



Şekil 8: Güvenilirlik katsayısı

Güvenilirlik, genellikle birden çok uygulamaya dayalı yöntemler ve tek uygulamaya dayalı yöntemler başlıkları altında ele alınır. Birden çok uygulamaya dayalı yöntemler altında test-tekrar test ve eşdeğer (paralel) testler yöntemleri, tek uygulamaya dayalı yöntemler altında ise eşdeğer yarılar, KR-20, KR-21, Cronbach alfa, Hoyt'un varyans analizi, McDonald omega vb. yöntemler yer almaktadır.





Şekil 9: Güvenilirlik kestirim yöntemleri

2.5.1. TEST-TEKRAR TEST YÖNTEMİ: Bu yöntem ile test güvenilirliğini test etmek için bir test, aynı gruba, belli bir zaman aralığıyla iki kez uygulanır. Daha sonra bireylerin birinci uygulamadan aldıkları puanlarla ikinci uygulamadan aldıkları puanlar arasındaki korelasyon hesaplanır. Elde edilen korelasyon katsayısına **kararlılık (devamlılık / istikrarlılık) katsayısı** adı verilir.

	Anksiyete Ölçeği Puanları X_1	Anksiyete Ölçeği Puanları X_2
Ceren	54	50
Filiz	98	110
Aydın	128	120
Ece	89	80
Deniz	101	101
....	87	85
....	98	106
....	67	59
....	122	119
Dursun	79	82

Kararlılık Katsayısı

Korelasyon (r) 1'e yakın

Şekil 10: Test-tekrar test yöntemi çalışma örneği

Güvenilirliğin kararlılık, tutarlılık ve duyarlılık sorunu olduğuna daha önce değinilmişti. Bu yöntemle elde edilen güvenilirlik kanıtı, güvenilirliğin yalnızca kararlılık boyutuna karşılık gelir; güvenilirliğe ilişkin tümel bir sonuç ortaya koymaz. Bu nedenle bir ölçme aracının güvenilirliğine ilişkin kanıt toplamak isteyen bir kişi kararlılığın yanı sıra tutarlılığa ve duyarlılığa ilişkin sorgulamaları yapmak durumundadır.

Bu yöntem, ölçülen özelliğin kararlı olduğu durumlarda uygulanması gereken bir yöntemdir. Testin ölçtüğü özellik sürekli değişkenlik gösteriyorsa bu yöntemle testin güvenirliği hesaplanmamalıdır. Diğer bir deyişle bu yöntem daha çok iki uygulama arasında kolaylıkla değişmeyen özellikleri ölçen testler için uygundur. Örneğin genel zihin yetenekleri, kişilik testleri, ilgi envanterleri, tutum ölçekleri vb. gibi testlerin kararlılık bağlamında güvenilirlikleri bu yöntemle hesaplanabilir.

Bu yolla test güvenilirliğini kestirmede karşılaşılan önemli bir sorun, testin iki uygulanışı arasındaki zaman aralığının ne kadar olması gerektiğidir. Bu soruya kesin bir yanıt vermek olanaksızdır. Aradan geçen zaman, hem ölçme aracının ölçtüğü özellik bakımından



yanıtlayıcıların önemli ölçüde değişmelerine hem de birinci uygulamada verilmiş olan yanıtların hatırlanmasına izin vermeyecek bir uzunlukta olmalıdır. Diğer bir ifadeyle iki uygulama arasındaki zamanın belirlenmesindeki temel ilke şudur: *"Birinci uygulamada maddeleri hatırlamayacak kadar uzun, ölçülen özellikte köklü değişimler oluşmayacak kadar kısa olmalıdır."*

Aradaki zamanın belirlenmesinde dikkat edilmesi gereken bir başka nokta, özelliğin değişim hızıdır. Bazı özellikler daha hızlı değişime açıkken bazıları ise daha zor değişir. Örneğin bireylerin genel yetenek düzeylerinin kısa bir zamanda değişmesi pek olanaklı değilken tutumu görece daha kolay değişebilir. Dolayısıyla bir genel yetenek testi için aradaki zaman daha uzun tutulabilirken tutum ölçeği için daha kısa tutulabilir.

Ayrıca özelliklerin değişimi yaşla da ilişkilidir. Bazı özellikler belli yaşlarda daha durağanken bazı yaşlarda daha değişkendir. Örneğin yaşamın ilk evrelerinde çocukların gelişimi daha hızlı iken yaşın ilerlemesi ile birlikte gelişim yavaşlayabilir. Dolayısıyla bebeklik döneminde uygulanacak bir gelişim envanteri için sürenin daha kısa tutulması gerekebilir.

Test-tekrar test yöntemi, başarı testleri için çok tercih edilmemektedir. Çünkü kısa vadede başarı kolay değişebilen, kararsızlık gösterebilen bir özelliktir. Bu yöntem daha çok yetenek testleri, kişilik envanterleri vb. psikolojik ölçme araçlarından elde edilen puanların güvenilirlik kanıtlarını üretmek için tercih edilmektedir.

Birden çok uygulamaya dayalı yöntemler alanyazında genellikle benzer ölçekler geçerliliği, hâlihazır geçerlilik, uygunluk geçerliliği olarak da geçer.

Birden çok uygulamaya dayalı yöntemlerle güvenilirlik kanıtı elde etmek biraz zahmetlidir. Test-tekrar test yönteminde testi farklı zamanda ikinci kez uygulamak için grup bulmak bazen zor olabilir. Ayrıca katılımcı kaybı yaşamak da olasıdır. Tek uygulamaya dayalı yöntemler tek bir test, tek bir grup ve tek bir uygulama gerektirmektedir. Dolayısıyla güvenilirlik kanıtı elde etmek daha pratiktir (Özellikle kararlılık katsayısının hesaplanmadığı durumlarda güvenilirliğin bir boyutu tabii ki eksik kalır.). Bir test bir gruba, bir kez uygulanıyorsa test kendi içinde değerlendirilecek demektir.

Güvenilirliğin bir boyutu da testin tutarlılığıdır. Tek uygulamaya dayalı güvenilirlik sorgulama yöntemleri ile "Test kendi içinde tutarlı bir bütün oluşturur mu?" sorusuna yanıt aranır. Bu nedenle bu yöntemlerin tümünden elde edilen katsayı **iç tutarlılık katsayısı** olarak adlandırılır.

2.5.2. TEST YARILAMA (EŞDEĞER YARILAR / İKİ YARI GÜVENİLİRLİĞİ) YÖNTEMİ: Bu yöntemle güvenilirliği tahmin etmede uygulanmış bir test iki eşdeğer yarıya bölünür ve bireylerin iki yarıdan aldıkları puanlar arasındaki tutarlılık incelenir. Buradaki en temel sorunlardan biri testin iki eşdeğer yarıya nasıl bölünmesi gerektiği ile ilgilidir. En sık başvurulan yöntemler: a) ilk yarı ve son yarı b) tek ve çift ve c) rastlantısaldır. İlk ve son yarı yöntemi her test için uygun değildir. Uygun olmadığı durumlar: Testteki

- maddeler basitten zora doğru sıralanmış ise
- maddeler konu içeriklerine ya da faktörlere göre kümelenerek yerleştirilmiş ise
- madde sayısı çok fazla ise

	Türkçe Testi	$X_{1.YARI}$	$X_{2.YARI}$
Deniz	48	20	28
Şebnem	98	50	48
Ufuk	66	32	34
Işıl	88	44	44
Sinem	32	20	12
……	70	30	40
……	62	28	34
……	40	20	20
……	60	28	32
Sinan	66	36	30

Testin yarısına ilişkin tutarlılığı verir. Bu korelasyon testin tamamına nasıl genellenir?

Korelasyon (r)

Spearman-Brown Eşitliği

İç tutarlılık katsayısı

Şekil 11: Test yarılama yöntemi çalışma örneği

Bu yöntemle güvenilirliği tahmin etmede, uygulanmış bir test iki eşdeğer yarıya bölünür ve bireylerin iki yarıdan aldıkları puanlar arasındaki tutarlılık incelenir. Şekil 11'de de görüldüğü üzere iki eşdeğer yarıdan alınan toplam puanlar arasında korelasyon hesaplanır ancak elde edilen katsayı korelasyonun yarı puanlar üzerinden elde edilmesi nedeniyle testin tamamına ilişkin bir tutarlılık bilgisi vermez, yarısına ilişkin verir. Diğer bir ifadeyle elde edilen katsayı eşdeğer yarılardan birinin güvenilirliği olarak kabul edilir. Testin tümüne ilişkin bir güvenilirlik katsayısı Spearman-Brown formülü aracılığıyla hesaplanır. Elde edilen katsayı testin tamamına ilişkin **iç tutarlılık** bağlamındaki güvenilirlik katsayısı olarak kabul edilir.

2.5.3. KUDER-RICHARDSON 20 VE 21 YÖNTEMLERİ: Kuder-Richarson 20 ve 21 ya da kısaca KR-20 ve KR-21, testin kendi içinde tutarlı bir bütün oluşturup oluşturmadığı ile ilgili bilgi verir. Bu nedenle her iki yöntemle de testin iç tutarlılığı değerlendirildiğinden bu yöntemlerden elde edilen katsayıya **iç tutarlılık katsayısı** adı verilir.

KR-20 ve KR-21 ile güvenilirlik kestirimi, puanlamanın kategorik olduğu ölçme araçları için uygundur. Puanlamanın kategorik olması, doğru yanıtlara 1 puan, yanlış ve boş bırakılan maddelere 0 puan vererek puanlamanın yapıldığı ya da maddenin yanıtının iki seçenekli verildiği durumlardır. Eğer testteki maddeler farklı ağırlıklarla puanlanmışsa ya da test puanları şans başarısı için düzeltilmişse bu formüller kullanılmaz.

2.5.4. CRONBACH ALFA YÖNTEMİ: Eğer ölçme aracının puanlaması çok kategorili ise KR-20 ile aynı mantık üzerine kurulu Cronbach alfa hesaplanır. Elde edilen katsayının ismi yine **iç tutarlılık katsayısı**dır. Derecelendirme ölçekleri puanlamanın çok kategorili olduğu araçlara örnek verilebilir.

2.6. GÜVENİLİRLİĞİ VE GEÇERLİLİĞİ ARTIRMA YOLLARI:

Güvenilirlik temelde tesadüfi hatalardan arınık olma durumu ise şüphesiz ki tesadüfi hata kaynaklarını azaltmaya yönelik eylemler güvenilirliği artıracaktır. Ancak bu noktada güvenilirliği artıran temel kavramlar ve eylemler konusunun ayrıntılarına girelim:

1. Bir testteki madde sayısı arttıkça birimler küçüldüğünden duyarlılık artar, bu nedenle hata miktarı azalacağından güvenilirlik artar.
2. Bir ölçme işleminde genel bir ilke olarak puanlayıcı sayısı arttıkça güvenilirlik artar. Birden çok kişinin yaptığı puanlamanın ortalamasını almak, bir kişinin verdiği puandan daha güvenilir olma eğilimindedir.



3. Puanlama nesnelliği, güvenilirliği büyük ölçüde etkiler. Bir testin farklı kişiler tarafından puanlanması ya da aynı kişinin farklı zamanlarda verdiği puanlar arasındaki tutarlılığa **puanlama güvenilirliği** adı verilir. Bir testten elde edilen puan, puanlayıcıya ya da zamana göre değişmiyorsa testin güvenilirliği artar. Elbette kullanılan ölçme aracının puanlamasının nesnel ve öznel olması bu noktada çok önemlidir. Çoktan seçmeli, derecelendirme ölçeği, kontrol listeleri gibi araçlar puanlaması nesnel araçlara örnek iken açık uçlu sorulardan oluşan araçlar, mülakatlar vb. ise puanlaması öznel araçlara örnek olarak verilebilir. Öznel araçların puanlama biçiminin mümkün olduğunca standartlaştırılarak nesnel hâle getirilmesi güvenilirlik açısından önemlidir.
4. KR-20 ve KR-21 yöntemlerinde bahsedildiği gibi testteki maddeler açısından benzeşiklik (homojenlik) arttıkça güvenilirliğin artması, diğer taraftan ayrışıklık (heterojenlik) arttıkça güvenilirliğin düşmesi beklenen bir durumdur.
5. Testten elde edilen puanların güvenilirlik kestirimi için veri elde edilecek grubun (örneklem) büyüklüğü arttıkça grubun heterojenleşmesinden dolayı güvenilirlik artar.
6. Aslında bir üst maddeyle de paralel bir biçimde maksimum performansı ölçen testlerde ortalama güçlüğe yaklaştıkça ($\bar{P} = .50$) grup heterojenleşir ve dolayısıyla güvenilirlik artar. Heterojen gruplarda testin ölçmeye konu olan özelliğe sahip olanla olmayanı ayırt etme gücünün artması beklenir. Grubun ölçülen özellik bağlamında heterojenleşmesi, büyük ölçüde örneklem sayısının artmasına ve eğer bilişsel bir özellik ölçülüyorsa testin ortalama güçlüğe yaklaşmasına bağlıdır.
7. Maddelerin dil bilgisi kurallarına uygun, açık ve anlaşılır yazılması belki de güvenilirliği en çok artıran etmenlerden biridir. Bu koşul sağlanmadığı takdirde diğer koşulların çok fazla önemi olmayabilir. İyi madde yazarı olmanın temel özelliklerinden biri de ana dilini iyi kullanma becerisidir. Dolayısıyla test yazar(lar)ının dil bilgisi ve imla kurallarına uygun bir biçimde maddeleri yazması, maddede yoklamak istediklerini yalın ve kısa bir biçimde ifade etmesi gerekir. İdealde bu konudaki temel ilke, bir maddenin tüm öğrenciler tarafından aynı biçimde anlaşılmasıdır.
8. Öğrencilere test uygulaması öncesinde ve gerekliyse süreçte yönerge vermek, test almaya güdülenme ve hazırbulunuşluluğun artması açısından önemlidir. Buna yönelik alınacak önlemler özellikle bireyden kaynaklanabilecek hataları asgariye indirir.
9. Fiziksel ortamın ses, sıcaklık, ışık, koku, dikkat dağıtıcı uyarıcılar, ergonomi vb. etmenler açısından uygun hâle getirilmesi yine güvenilirliği artıran bir başka etmendir. Fiziksel uyarıcılar açısından yanıtlama davranışını etkileyebilecek her türlü etmen güvenilirliği düşürecektir. Bu nedenle test uygulayıcılarının test ortamını düzenleme bağlamında sorumlulukları vardır. Özellikle zekâ, başarı gibi bilişsel özelliklerin ölçülmesinde bu etmenlerin önemi daha da artmaktadır.
10. Eğer süreli bir test uygulaması ise sürenin yeterli verilmesi gerekir. Eğer bir hız testi değilse tüm katılımcıların maddelerin tümüne ulaşması beklenen bir durumdur. Tabii ki hız testleri bu durumdan bağımsız olarak değerlendirilebilir.
11. Uygulamada bireylerin dikkatini dağıtacak ve/veya kaygısını artıracak davranışlardan kaçınılması, puanlama ya da veri girişi yapılırken dikkatli ve titiz davranılması ölçme işlemini yapan kişiden kaynaklanabilecek hataları azaltacağından dolayı güvenilirliği artırır.
12. Maddelerin teste düzgün yerleştirilmesi, okumayı güçleştirecek bir unsurun bulunmaması; test katılımcısının yaşına, gelişim düzeyine uygun bir punto büyüklüğü kullanılması; baskı hataları bulunmaması vb. etmenler güvenilirliği artırır.

Güvenilirliği artıran faktörler geçerliliği de artırır. Ancak geçerlilik yalnızca tesadüfi hatalardan değil, aynı zamanda sabit ve sistematik hatalardan da etkilenir. O nedenle öğretmenin



tesadüfi hataların yanı sıra sabit ve sistematik hatalardan arındırmaya yönelik alacağı tüm tedbirler geçerliliği de artıracaktır.

3.TEST GELİŞTİRME VE MADDE / SORU TÜRLERİ

Doç. Dr. Güçlü ŞEKERCİOĞLU

Test, eğitimde bireylerin özelliklerini belirlemeye yönelik ölçme araçlarına verilen genel bir kavramdır. Bu terim zaman zaman günlük dilde yalnızca çoktan seçmeli maddelerin bulunduğu ölçme araçları için kullanılsa da test bu belirlemeden çok daha geniş bir anlamda kullanılmaktadır.

3.1. TEST TÜRLERİ:

Tablo 4: Test türleri

Testi Alan Kişi Sayısına Göre	Testin Uygulanış Süresine Göre	Testin Ölçtüğü Niteliğe Göre	Değerlendirme Yaklaşımına Göre	Hazırlanış Biçimine Göre	Veri Toplama Tekniğine Göre
1. Bireysel	1. Süreli	1. Hız testleri	1. Objektif	1. Standart	1.Performans
2. Grup	2. Süresiz	2. Güç Testleri	2. Subjektif	2. Öğretmen yapımı	2. Kağıt-Kalem

3.2. TEST GELİŞTİRME: Önceden belirlenmiş özelliklerin ölçülmesi ile ilgili bir test geliştirilmesi, sistematik işlem basamaklarına uyulmasını gerektirir. Test geliştirmede farklı yaklaşımlar olmakla birlikte burada, alanda yaygın kabul gören işlem basamakları üzerinde durulacaktır. Test geliştirmenin işlem basamakları: a. Amacın belirlenmesi, b. Kapsamın belirlenmesi ve belirtke tablosunun oluşturulması, c. Denemelik maddelerin / soruların yazılması, d. Maddelerin / soruların gözden geçirilmesi (redaksiyon), e. Denemelik test formunun hazırlanması, f. Testin uygulanması, g. Test ve madde istatistiklerinin hesaplanması, h. Seçilen maddelerden oluşan nihai formun oluşturulması.

Pek çok kaynakta test geliştirmenin adımları olarak bunlar olsa da sınıf içi ölçme ve değerlendirme etkinliklerinde bu basamaklara uygun test geliştirmek pratikte çok olanaklı değildir. Bu durumda öğretmenler açısından sınıf içi ölçme ve değerlendirmelerde uygulanması daha olanaklı bir test planı önerelim:

1. Sınavın amacının belirlenmesi (puanların hangi amaçlarla kullanılacağı)
2. Sınavda yoklanacak davranışların belirlenmesi
3. Sınavın kapsayacağı konuların listelenmesi
4. Okulun takviminin ve sınav sonuçlarının kullanılacağı zamanın dikkate alındığı bir sınav gününün belirlenmesi
5. Bir ders saatine uygulanabilecek uzunlukta bir sınav süresinin belirlenmesi,
6. Belirtke tablosunun hazırlanması
7. Soru/madde türlerinin belirlenmesi [Soru/madde türleri sınavda kullanılacak davranışlara göre belirlenir. Bir sınavda tek bir soru/madde türü kullanılabileceği gibi birden fazla da kullanılabilir (soru/madde çeşitlemesi).]
8. Soru/madde sayısının belirlenmesi
9. Sınav süresinin belirlenmesi



10. Sınavda kullanılacak soruların/maddelerin “ortalama güçlüğünün” ve “güçlük dağılımının” belirlenmesi
11. Soruların/maddelerin yazımında, redaksiyonunda ve teste alınacak soruların/maddelerin seçiminde izlenecek yolun belirlenmesi
12. Cevap anahtarının ve puanlama yönteminin belirlenmesi
13. Ölçme aracını yazma ve çoğaltma yönteminin belirlenmesi
14. Sınavın uygulanma kurallarının belirlenmesi (yönerge yazılması)
15. Sınavın uygulanması
16. Test ve madde istatistiklerinin hesaplanması

3.3. BAŞARININ ÖLÇÜLMESİNDE YÖNTEMLER



Şekil 12: Ölçme araç ve yöntemleri



Şekil 13: Geleneksel yöntemler (Madde / soru türleri)

3.3.1. GELENEKSEL ÖLÇME YÖNTEMLERİ:

3.3.1.1. DOĞRU-YANLIŞ MADDELERİ: Cevaplayıcının verilen ifadelerin doğru mu, yanlış mı olduğunu belirlemesinin istendiği madde türüdür. Doğruyanlış maddelerinde verilen bir tür “**önerme**”dir.



Cumhuriyet 1920 yılında kurulmuştur. (D) (Y)

Çoktan seçmeli madde türünden sonra en çok kullanılan ikinci madde tipidir. Genellikle öğrencinin bilimsel gerçekleri, tarihî olayları, kesin yargıları tanıması ve hatırlaması; olguları ve gerçekleri kişisel yargı ve görüşlerden ayırt etmesi ölçülmek istendiğinde kullanışlıdır.

Avantajları: a. Hazırlaması, uygulaması ve puanlaması kolay, diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematik hata karışma olasılığı pek yoktur.

Dezavantajları: a. Şans başarısı olasılığı yüksektir. b. İyi yazılmazsa genellikle üst düzey bilişsel özellikleri ölçemez. c. Öğrenme eksiklerini belirleyemez. d. Yanlış üzerinden öğretim yapılmaz. e. Her derste/konuda kesin yanlış denebilecek durumlar belirlemek zordur.

Doğru-Yanlış Maddeleri Yazım Kuralları:

- Her madde tek ve belirli bir fikri belirtmelidir. Özellikle aynı maddede biri doğru, öteki yanlış iki fikir ifade etmekten kaçınılmalıdır.
- Doğru yanlış maddesi, kesinlikle doğru ya da kesinlikle yanlış olmalıdır. Maddenin doğruluğu ya da yanlışlığı, başka bir açıklamaya gerek kalmadan belirlenebilmelidir.
- Bir maddenin yanlışlığı önemsiz bir ayrıntıda ya da aldatıcı bir noktada olmamalıdır. Bir yargı temelden yanlış olmalıdır.
- Mümkün olduğunca olumsuz ifade kullanılmamalıdır. Özellikle iki olumsuz ifadeden kesinlikle kaçınılmalıdır.
- Kanı ifadeleri kendiliklerinden doğru ya da yanlış olmadığından bir kaynağa dayandırılarak verilmelidir. Belli bir kaynağa ya da bir otoriteye dayandırılmadan verilen kanı ifadelerinin yer aldığı maddelerin ayrıt etme gücü sıfır ya da negatif olma eğilimindedir.
- Bir maddenin ifadesi kısa, açık ve yalın olmalıdır. Madde, onda sorulan ana fikrin doğruluk ya da yanlışlığının belirlenmesi için gerekli olmayan ayrıntılarla şişirilerek uzatılmamalıdır.
- Okuduğunu anlama süreçlerinin ilk basamağı, metin içerisinde açıkça verilmiş bilgiden doğrudan çıkarım yapmadır. Bir metne bağlı olarak yanıtlanan doğru-yanlış maddelerinin anlam geliştirme, yorumlama, değerlendirme gibi daha üst düzey okuduğunu anlama süreçlerine yönelik olabilmesi için, metindeki cümleler aynen yazılmamalıdır.
- Öğrencinin bir maddede ifadenin doğruluğuna ya da yanlışlığına karar vermesi gerektiğinde öğretmen özellikle vurgulamak istediği bir sözcük / terim vs. varsa onun altını çizebilir.
- Doğru-yanlış maddesinin özellikle yanlış olduğu durumlarda, öğrencinin şansla puan almasını engellemek amacıyla ifadeyi düzeltmesi ya da yanlışlığın ne olduğunu belirtmesi istenebilir.
- İyi kurgulandığı takdirde doğru-yanlış maddeleri ile üst düzey öğrenmeler de yoklanabilir.
- Doğru-yanlış maddeleri aynı kapsamdan geliyorsa ortak bir soru kökü altında birleştirilebilir.
- Doğru ve yanlış maddeler, testte belirli bir örüntüye göre yerleştirilmelidir. Maddelerin sıralanışı, baştan itibaren iki doğru bir yanlış, bir doğru iki yanlış gibi bir örüntü gösterirse öğrencilerin bunu fark etmeleri olanaklı olabilir. Bu nedenle doğru ve yanlış maddelerin testteki sıralanışı rastgele olmalıdır.



- Doğru ve yanlış maddelerin ifadesi yaklaşık olarak aynı uzunlukta olmalıdır.
- Bir doğru-yanlış testindeki doğru ve yanlış maddelerin sayısı, yaklaşık olarak birbirine eşit olmalıdır.
- Maddeleri işaretleme yöntemi, öğrenciye açık ve anlaşılır bir biçimde açıklanmalıdır.
- Mümkün olduğunca kısa yazılmalı, basit cümle yapısı tercih edilmeli, özellikle "ve" gibi bağlaçlara dikkat edilmelidir.
- Olumsuz anlama sahip bir sözcük ya da ifade varsa altı çizilmelidir.
- İpucu vermekten kaçınmak gerekir. Özellikle "asla, daima, hiçbiri, tümü" vb. ifadelerin yanlış olma olasılığı yüksektir. Öte yandan "genellikle, çoğunlukla, bazen, zaman zaman, sıklıkla" vb. ifadelerin de doğru olma olasılığı yine yüksektir.
- Üst düzey becerileri yoklamak için tablo, grafik, harita ya da okuma parçası gibi bir öncülün kullanılması önerilir.

3.3.1.2. EŞLEŞTİRME MADDELERİ:

Eşleştirme maddeleri, iki grup hâlinde verilen ve birbirleriyle ilgili olan bilgi ögelerinin belli bir açıklamaya göre eşleştirilmesini gerektirir.

Yönerge / Açıklama

Yönerge: Aşağıdaki "A" sütununda Türkiye Cumhuriyeti tarihi ile ilgili bazı önemli olaylar, "B" sütununda ise bu olayların geçtiği yıllar verilmiştir. Bir olayın hangi yılda geçtiğini kararlaştırdıktan sonra, o yılın önündeki harfi, olayın solundaki boşluğa yazınız. "B" sütunundaki bir tarih, bir veya birden fazla kez kullanılabileceği gibi, hiç kullanılmayabilir.

Öncüller / Maddeler | **Muhtemel cevaplar / Seçenekler**

	A Sütunu	B Sütunu
C	1. Osmanlı Meclis-i Mebusan'ın son toplanması	A. 1918
D	2. Ankara'nın başkent olması	B. 1919
H	3. Tek dereceli seçimlerin ilk kez yapılması	C. 1923
C	4. TBMM'nin açılması	D. 1925
D	5. Lozan Antlaşmasının imzalanması	E. 1928
D	6. Cumhuriyet'in ilanı	F. 1930
C	7. Sevr Antlaşmasının imzalanması	G. 1934
		H. 1936
		I. 1946

Avantajları: a. Hazırlaması, uygulaması ve puanlaması kolay, diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematik hata karışma olasılığı pek yoktur.

Dezavantajları: a. Şans başarısı olasılığı yüksektir. b. İyi yazılmazsa genellikle üst düzey bilişsel özellikleri ölçemez.

Eşleştirme Maddeleri Yazım Kuralları:

- Her eşleştirme soru grubunda yönerge/açıklama yazılmalıdır. Yönergede eşleştirmeye konu olan kapsam belirtilmeli, öncüllere ve seçeneklere nelerin konulduğuna değinilmelidir. Eşleştirmenin nasıl yapılacağı konusunda bilgi verilmelidir.
- Seçeneklerin kullanılıp kullanılmayacağı, birden fazla kez kullanma durumu belirtilmelidir. Bir eşleşme takımındaki öncüller ile cevaplar eşit sayıda olmamalıdır. Eğer eşit sayıda olursa öğrenci, hemen ilk bakışta bildiklerini eşleştirecek; geri kalan öncüllerin cevaplarını da tahmin edecektir.
- Eşleştirmede ortaokul ve üstü öğretim kademelerinde öncülün başında bir boşluk bırakılarak öğrenciden seçeneğin harfini yazmaları istenebilir. İlkokul düzeyinde ise öğrencilerden öncüllerle seçenekler arasında oklarla eşleştirme yapmaları istenebilir.
- Okul öncesinde, özel eğitimde eşleştirme şekillerle ya da resimlerle / fotoğraflarla yapılabilir.



- Bir eşleştirme maddesinde yer alan öncüller ile seçeneklerin her biri benzeşik (homojen) ögelerden oluşmalıdır. Diğer bir ifadeyle aynı kapsamdan gelmelidir.
- Uzun ifadeler öncül olarak kullanılmalı ve öncüller sütunu sayfanın sol sütununa, cevaplar sütunu ise sayfanın sağ tarafına yerleştirilmelidir. Böyle olması, cevaplamada zaman kaybını önler ve cevabın seçimini kolaylaştırır.
- Bir eşleştirmeli maddeler grubundaki madde sayısı, en az 6, en çok 15 olmalıdır. Madde sayısının 6'dan az olması, salt tahminle doğru cevabın bulunma olasılığını artırır. 15'ten çok madde kullanıldığında ise cevaplayıcıların doğru olarak eşleştirilecek ifadeleri ayıklayıp seçmesi çok zaman alır. Üstelik madde sayısı arttıkça öncüller ile cevaplar takımının benzeşikliğini sağlamak giderek güçleşir.
- Maddelerin tümü aynı sayfada bulunmalıdır. Aksi hâlde cevabın seçimini güçleştirir.
- Cevapların seçileceği sütun, bir kelime listesi ise alfabetik sıraya göre (küçükten büyüğe ya da büyükten küçüğe) düzenlenmelidir.
- Bu eşleştirme maddeleri öğrencilerin yalnızca hatırlama düzeyi dışında, kavrama düzeyindeki öğrenmelerinin de yoklanabileceğine ilişkin örnektir. Örneğin grafik / tablo / metin / harita

3.3.1.3. ÇOKTAN SEÇMELİ MADDELER:

Soru 24 BÖLGELER ARASI BAĞLANTILAR

Madde kökü

	A	B	C	D
A	0	0	1	1
B	0	0	1	1
C	1	1	0	1
D	1	1	1	0

Öncül

Yukarıda verilen tablo aşağıdaki şekillerden hangisine ait olabilir?

Doğru yanıt

A) A, C, D, B
B) A, B, D, C
C) A, C, D, B
D) C, D, A, B

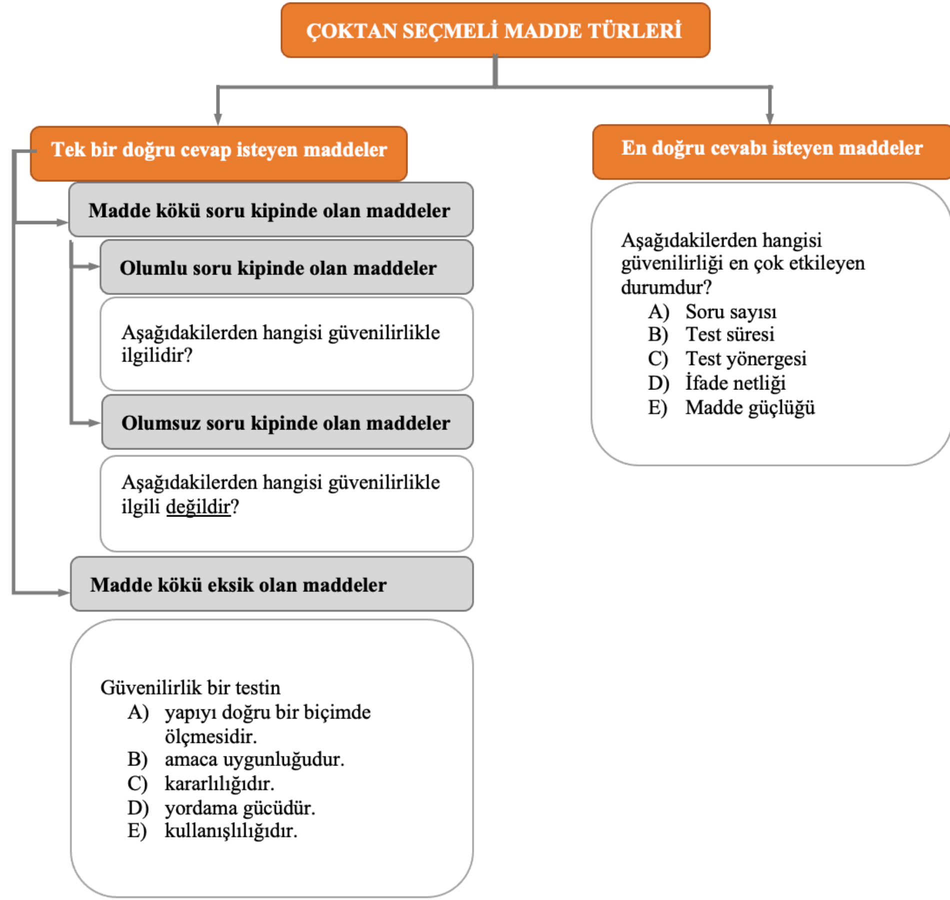
Çoktan seçmeli madde

Seçenekler

Çoktan Seçmeli Maddelerin Özellikleri:

- Hazırlanması zor ancak puanlaması kolaydır.
- Puanlama objektiftir.
- Üst düzey bilişsel becerilerle ilgili madde yazmak zordur. O yüzden yazılan maddeler genellikle bilgi, kavrama ve uygulama basamaklarındadır.
- Doğru cevabı içinde barındırması nedeniyle sentez düzeyinde soru sormak olanaklı değildir. Bu nedenle bir test yalnızca çoktan seçmeli maddelerden oluşuyorsa yaratıcılığın denetlenmediği yönünde haklı bir eleştiri vardır.
- Doğru cevabı içinde barındırması nedeniyle puanlara şans başarısı karışma ihtimali vardır.
- Çok sayıda soru sorulabilir.
- Sistematik hata karışma ihtimali çok düşüktür.
- Çeldirme mantığı temellidir.





Şekil 14: Çoktan seçmeli madde türleri

Çoktan Seçmeli Madde Yazım Kuralları:

- İdeal bir çoktan seçmeli maddede, bilen bir öğrenci, seçenekleri görmesine gerek kalmaksızın doğru cevabı verebilmelidir.
- Öğrencinin sözcük repertuvarında bulunmayan, henüz öğrenmediği sözcük, kavram, terim vb. kullanılmamalıdır.
- Bir durumun resim ya da fotoğrafla anlatılabildiği durumda, çokça cümle kurmak yerine görsellerden yararlanılabilir. Ancak maddenin cinsiyet gibi alt gruplarda yanlılık oluşturmamasına dikkat edilmelidir.
- Gereksiz görsel kullanılmamasına da özen gösterilmelidir. Görsel çıkarıldığında anlamca bir kayıp oluşmuyorsa o görsel kullanılmamalıdır.
- Bir maddeyi okuyan bir öğrenci öğretmenin zihninden geçeni keşfetmek zorunda değildir.
- Bazı öğretmenler karışık bir dille madde yazdığında sorunun zorlaşacağını düşünür. Oysa bu oldukça yanlış bir yöntemdir. Karışık bir dil yerine ortalama bir öğrencinin anlayabileceği bir dil kullanılmalıdır. Bir çoktan seçmeli maddeyi zorlaştırmak isteyen öğretmen temelde iki yöntem kullanabilir: a. Çeldiricileri doğru yanıta yaklaştırmak (güçlendirmek) ve b. Davranışı üst düzey becerileri ölçen bir formda hazırlamak.
- Madde kökünde gereksiz yere sözcük kullanılmamalıdır. Zaman zaman öğretmenler kendi doğrularını ya da önemli yerleri vurgularken bu hatayı yapmaktadırlar.
- Seçeneklerde gereksiz sözcük tekrarından kaçınılmalıdır. Tekrar edilen sözcük madde köküne alınabilir.



- Bir maddenin doğru yanıtı, başka bir maddenin kökünde ya da seçeneklerinde bulunmamalıdır.
- Seçeneklerin hepsi birbirleriyle tutarlı ve ayrı gramatik yapıya sahip olmalıdır. Eğer seçenekler ifade bakımından birbirinden farklıysa çoğu durumda bu öğrencilere ipucu veren bir yapı sergiler. Özellikle eklerin kullanımına dikkat etmek gerekir.
- İpucu vermekten kaçınmak gerekir. Özellikle "bazen, çoğunlukla, genellikle, sıklıkla ya da zaman zaman" vb. ifadelerin doğru olma olasılığı yüksektir. Diğer taraftan "hiçbir zaman, asla, daima, hiçbiri, her zaman, tümü" vb. ifadelerin de yanlış olma olasılığı yüksektir. Bu nedenle bu tür sözcükler seçeneklerde kullanılmamalıdır.
- Birbirinin tam zıttı olan durumların seçeneklerde verilmesi çoğunlukla uygun değildir. Çünkü genellikle ikisinden biri doğrudur. Şans başarısı artar, geçerlilik ve güvenilirlik düşer.
- Seçenekler yazılırken anlamca birbirini içeren / kapsayan ifadelerin kullanılmamasına dikkat edilmelidir.
- Seçeneklerin uzunlukları birbirine eşit olmalıdır. Özellikle doğru cevabın daha uzun ya da daha kısa olmamasına özen gösterilmelidir.
- Doğru cevaplar teste dağıtılırken dikkat edilmeli, doğru cevapların bir örüntü göstermemesi sağlanmalıdır.
- "Hepsi" ve "hiçbiri" bir seçenek olarak kullanılmamalıdır.
- Çoktan seçmeli maddelerde seçenek sayısı genellikle 4 ya da 5 olsa da seçenek sayısını belirleyen temel faktör öğrencinin içinde olduğu gelişim dönemidir. Lise ve daha üstü öğretim düzeylerinde 5, ortaokul düzeyinde 4 ve ilkokul düzeyinde 3 seçenek kullanılabilir.
- Seçenekler bir sıra ile verilmelidir. Seçenekler baş harflerine göre alfabetik olarak, eğer sayı iseler büyüklük olarak sıralanmalıdır.
- Hangi özellik ölçülürse ölçülsün, çoktan seçmeli madde yazımının bütün aşamalarında hem madde kökünde hem de seçeneklerde dil bilgisi ve imla kuralları konusunda yüksek bir titizlik ve özen göstermek çok önemlidir. İyi bir madde yazarı; a. madde yazılan alanı çok iyi bilmeli, b. madde yazma teknik ve yöntemlerinden haberdar olmalı, c. maddelerin yazılacağı dili kullanmada becerikli olmalı ve d. testin uygulanacağı öğrencilerin gelişim düzeyini çok iyi bilmelidir.
- Çoktan seçmeli maddelerde seçeneklerin baş harfleri büyük olmalıdır.
- Bir madde kökü ile seçenekleri aynı sayfada olmalıdır. Yarısı başka sayfada, yarısı diğer sayfada olmamalıdır. Bu durum ortak köke dayalı maddeler için de geçerlidir.
- Eğer ortak köke dayalı madde grubu yazılacaksa hangi soruların ortak köke göre yanıtlanacağı belirtilmelidir. İlkokulda madde sayısını çok artırmamak gerekir.

AÇIK UÇLU SORULAR

Açık uçlu sorular ölçme / psikometri alanyazınında genellikle uzun yanıt gerektiren maddeler ve yanıtı sınırlı maddeler olarak ikiye ayrılır. Ancak bu testler sahada genellikle essay (yazılı yoklama), kısa yanıtlı, boşluk doldurma (cümle tamamlama) ve sözlü yoklama olarak geçmekte.

3.3.1.4. ESSAY (YAZILI YOKLAMA): Yazılı yoklamaların temel özellikleri:

- Cevaplayıcıların sorulara cevap verme konusunda sınırsız bir özgürlüğü vardır. Sınırsız cevap özgürlüğünün hem avantajı hem de dezavantajı bulunmaktadır.

  Avantajı: Öğrencinin yalnızca sahip olduğu bilgiyi değil, aynı zamanda zihninin nasıl işlediğini anlama olanağı sağlar.

  Dezavantajı: Öğrenci sınırlı bir bilgiye sahipse soruyla ilgisi olmayan, aklına estiği gibi uzun cevaplar yazabilir.



- Sınırsız cevap özgürlüğü, anlatım becerisi iyi olan öğrencilere bir avantaj sağlayabilir.
- Bu sınav türü, dünyada en eski ve günümüzde de öğretmenlerin hâlâ en çok tercih ettiği testlerden biridir. Bunun iki nedeni vardır: a. Pratik nedenler: Hazırlaması kolaydır. b. Eğitsel nedenler: Üst düzey bilişsel becerileri ölçmek için geleneksel yöntemler arasında en avantajlı yazılı test etme yoludur (Sentezde tek yol.).
- Yazma, problem çözme, bilgileri organize etme, analiz vb. özgün bir ürün ortaya koyma (yaratıcı düşünme), eleştirel düşünme, yeni durumlara beceriyi transfer etme, analitik düşünme, bilimsel düşünme, hipotez oluşturma ve neden-sonuç ilişkilerini açıklama, veri düzenleme, güçlü ve zayıf yönleri belirleme gibi becerileri ölçmek için çok kullanışlıdır.
- Çok fazla soru sorulamaması (Okuma ve yazma eylemlerinin süreler arasında ortalama 10 kat fark bulunmakta.) bir dezavantajıdır.
- Puanlar sadece öğrencinin sahip olduğu bilgiyi değil; öğrencinin anlatım biçimini, yazı güzelliğini, bilgisini örgütleme biçimini de yansıtır. Bu ise ölçme sonuçlarındaki sistematik hata miktarını artırabilir.
- Puanlama öznelliği belki de en önemli dezavantajıdır.

"Essay"de Nesnelliği Artırma Yolları:

- Ayrıntılı bir cevap anahtarı çıkarılmalıdır. Öğrenci ne yaparsa kaç puan verileceğinin belirlenmesi gerekir.
- Bütüncül (holistik) ya da analitik rubrik hazırlanabilir.

3.3.1.5. KISA CEVAPLI SORULAR: Kısa cevaplı testler; öğrencinin bir sözcük, bir rakam, bir tarih ya da en çok bir cümle ile cevaplandırabileceği sorulardan oluşur. Kısa cevaplı testler, öğrencinin aklının estiği cevabı yazmasını büyük ölçüde engelleyecek bir özelliğe sahiptir. Bundan ötürü maddeleri cevaplandıran bir cevaplayıcının yazacağı cevaplar çok kısadır. İstenen cevapların kısa ve belirgin olması, kısa cevaplı testlerin puanlanmasını genellikle nesnel kılar. Kısa cevap, soru cümlesi ya da eksik cümle yapısında olabileceği gibi hatırlama ya da tanımayı gerektiren bir yapıda da olabilir.

- Eğer konu kapsamı bakımında homojen bir yapıda ise gruplandırılarak sorulması daha uygun olacaktır.
- Uzun cevap gerektiren sorularda olduğu gibi bu soru türünde de ayrıntılı bir cevap anahtarı hazırlanmalıdır. Öğrencilerin verecekleri cevap çeşitliliğinin öğretmenin cevap repertuvarından her zaman daha fazla olacağı göz önüne alındığında, anahtarın zümrece hazırlanması önerilebilir.

Avantajları: a. Hazırlaması, uygulaması ve puanlaması kolay; diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematik hata karışma olasılığı pek yoktur.

İyi yazılmazsa genellikle üst düzey bilişsel özelliklerin ölçülememesi yine bir dezavantajıdır.

3.3.1.6. CÜMLE TAMAMLAMA SORULARI:
Cümle tamamlama, kısa cevaplı / sınırlı cevap gerektiren soru grubudur. Bu tür sorularda öğretmen önemli gördüğü bir cümleyi alır, kritik gördüğü bir ya da birden fazla yeri çıkarır ve yerine bir boşluk koyar. Öğrenciden o cümleyi anlamlı bir biçimde tamamlaması istenir. Cümle tamamlama soruları boşluk doldurma olarak da geçmektedir.

- Hazırlanması oldukça kolaydır. Bu yargı, özellikle *kim, ne, nerede, ne zaman?* sorularına cevap olabilecek olgusal bilgileri ölçmeye yönelik soruların yazılmasında geçerlidir.



- Cümleler ders kitabından aynen alınmamalı, öğretmen kendi sözcükleriyle soruları yeniden yazmalıdır.

Avantajları: a. Hazırlaması, uygulaması ve puanlaması kolay; diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematik hata karışma olasılığı pek yoktur.

İyi yazılmazsa genellikle üst düzey bilişsel özelliklerin ölçülememesi yine bir dezavantajıdır.

Cümle Tamamlama Soruları Yazım Kuralları:

- Her kısa cevap maddesi, yalnızca tek bir doğru cevabı olacak biçimde yapılandırılmalıdır. Tartışma götürecek, yoruma açık durumların soru yapılmasından kaçınılması ve sorunun ifadesinin açık, net ve anlaşılır olması gerekir.
- Bir maddenin ifadesinde, o maddenin cevabının bulunmasında işe yarayacak ipuçları vermekten kaçınılmalıdır. Doğru cevabın verilmesinde kullanılabilecek ipucu, maddede verilen gereksiz bir bilgi ya da cümlenin gramer yapısı olabilir.
- Cümlenin gramer yapısı bağlamında ipucu genellikle eklerle verilmektedir.
- Maddenin cevabında birden çok ayrıntı varsa o ayrıntıların her birine ayrı puan vermek gerekir.
- İlköğretimde cevapların ayrı bir cevap kâğıdına işaretletilmesi uygun görülmemektedir ancak ortaöğretimde bunun pek bir sakıncası bulunmamaktadır.
- Bir cümlede çok sayıda boşluk bırakılmamalı ve cümlelerden sadece anahtar niteliğindeki anlamlı ve önemli sözcükler çıkarılmalıdır. Eksik cümle istenileni anlatacak biçimde yapılanmış olmalıdır.
- Herkesçe aynı biçimde algılanacak, öğrencinin "Acaba öğretmenim bununla neyi kastediyor?" sorusunu sormayacağı biçimde soru yazmak gerekir.

3.3.1.7. SÖZLÜ YOKLAMA:

Sözlü sınavlar çok iyi organize edilse / kurgulansa dahi ciddi bir ölçme hatası barındırma potansiyeline sahiptir. Ancak bütün bu potansiyel sınırlılıklarına rağmen sözlü yoklamalar özellikle dile dayalı becerilerin ölçülmesi için geleneksel yöntemler içindeki tek türdür. Örneğin *sözlü anlatım, diksiyon, yabancı dil dersinde konuşma, solfej ve şarkı söyleme vb.*

- Sorular sözlü olarak sorulur ve cevap sözlü olarak verilir.
- Öğretmen ile öğrenci arasında karşılıklı ve devamlı bir etkileşim vardır.
- Bireysel bir test olması nedeniyle her öğrenciye ayrı soru sorma zorunluluğu vardır.
- Bireysel bir test olması nedeniyle tüm öğrencileri test etmek için gereken süre oldukça fazladır.
- Öğrencinin cevaplarını gözden geçirme şansı yoktur.
- Cevaplama çoğu zaman üzerinde düşünmeye ve tasarlamaya olanak olmadan verilir.
- Puanlama hemen yapılır. Bu ise cevabın doğruluğunun genel izlenimle yapılmasına neden olur.
- Öğrencinin yerinde ayağa kaldırılması ya da tahtaya kaldırılması kaygı vb. bireyden kaynaklanan hata miktarını artırabilir.
- Öğretmenin ölçme sonucuna sistematik hata karıştırma olasılığı yükselir.
- Sözlü anlatım becerisi iyi olan bir öğrenci, iyi olmayan bir başka öğrenciden daha az biliyor olsa da yüksek puan alabilir.
- Öğrencinin kılık-kıyafeti, diksiyonu, hâl ve hareketleri puanlamaya artı ya da eksi yönde etki edebilir.



- Çok fazla soru sorulamayacağı için kapsam geçerliliği problemi olabilir.

Sözlü Yoklama Uygulama Kuralları:

- Sözlü sınavların yapılacağı gün ve saat, yazılı sınavlarda olduğu gibi ilan edilmelidir.
- Sorular önceden hazırlanmalıdır. Soruların seçiminde amaca uygunluğa ve her öğrenciye benzer güçlükte soru belirlenmesine dikkat edilmelidir.
- Cevap anahtarı hazırlanmalıdır.
- Sınavın yapılması aşamasında öğrencilerin kaygısını artıracak davranışlardan kaçınılmalıdır. Hazırbulunuşluluk artırılmalıdır.
- Basit olan sorudan başlanmalıdır.
- Kayıt altına alın(a)mıyorsa puanlama hemen yapılmalıdır.
- Puana ilişkin dönüt hemen verilmelidir.
- Puanlamada sistematik hata kaynaklarına karşı dikkatli olunmalıdır.

Not: Yoklanacak olan davranışlar, yazılı sınav türleriyle de yoklanabiliyorsa sözlü yoklama yapmaktan kaçınılmalıdır.

GENEL DEĞERLENDİRME: Her madde / soru türünün avantajları olduğu gibi dezavantajları da vardır. Dezavantajları asgariye indirmenin yolu madde / soru çeşitlemesi yapmaktır ancak özellikle ilkokulda, çocukların gelişim süreçleri de dikkate alındığında, madde / soru çeşidi sayısını çok artırmamak gerekir.

Geleneksel test etme yaklaşımları ürün odaklıdır. Eğitim sürecinde ürün elbette önemlidir ancak ürün kadar öğrencinin süreçte bilgiyi nasıl yapılandırdığının izlenmesi de beklenen ürünün ortaya konulabilmesi açısından çok önemlidir. Bu nedenle geleneksel ölçme yöntemleri, destekleyici yöntemlerle (alternative assessment) desteklenmelidir.

3.3.2. DESTEKLEYİCİ DEĞERLENDİRME YAKLAŞIMLARI:

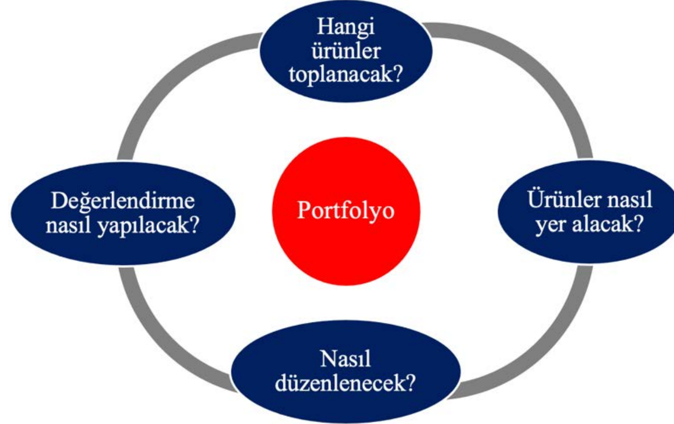
3.3.2.1. PORTFOLYO: Öğrencinin bir bütün olarak gelişim ve öğrenme süreci ile ürünlerini gösteren, aynı zamanda değerlendirilmesini de sağlayan sistemli ve amaçlı olarak oluşturulmuş dosyalardır. Portfolyolar ile amaçlanan, öğrenciyi çok yönlü olarak ve öğrencinin gelişimini bir süreç olarak izlemek ve değerlendirmektir. Portfolyolar ile öğrencinin öz geçmişi, ilgileri, tutumları, becerileri, yetenekleri, öğrenme biçimi, sosyal özellikleri vb. özelliklerinin tanınması amaçlanır. Sürecin paydaşları; öğrenci, öğretmen, arkadaş, aile ve iletişimde olduğu diğer kişilerdir. Öğrencinin yalnızca okulda değil; aile ortamında, oyun ortamında ve dış çevrede tanınması amaçlanır.

- Portfolyolar öğrencinin yaptığı çalışmaların bir araya getirildiği herhangi bir çalışma dosyası değildir.
- Öğrenciyi başarılı / başarısız olarak sınıflandırmayı sağlayan bir araç değildir.
- Öğrencilerin özelliklerini birbirleri ile karşılaştırmak amacıyla oluşturulmuş araçlar değildir.

Öğrenci portfolyosunun kabaca iki türü vardır. Bunlar:

a. Süreci yansıtan portfolyolar: Öğrencinin öğrenme ve gelişim sürecini yansıtır. Başlangıç çalışmalarını, süreçteki çalışmalarını, karşılaşılan güçlükleri ve öğrenme ürünlerini içerir.
b. Ürünü yansıtan portfolyolar: Öğrenme sürecinden çok bitmiş görevleri içerir. Öğrencinin en iyi olduğunu düşündüğü çalışmaları içerir.





Şekil 15: Portfolyo değerlendirme süreci

Portfolyonun içeriği:

- Öğretmen kayıtları (gözlemler, anektod kayıtları)
- Öğrencinin çalışmaları
- Öğrencinin sözel ve psikomotor becerilerini gösteren teyp ve video kayıtları
- Öğrencinin kendi çalışmaları hakkındaki düşünceleri, günlükler
- Öğrenciye yazılan mektuplar
- Öğrencinin yazdığı mektuplar
- Öğretmenin aileye ve diğer öğretmenlere yazdığı mektuplar

Portfolyo Oluşturma Süreci



TOPLAMA:

1. Hangi çalışmaların toplanacağına ve hangi özelliklerin gözleneceğine karar verilmesi
2. Öğrencilere çalışmalarının bir dosyada toplanacağının açıklanması ve çalışmalarını saklama konusunda öğrencilerin teşvik edilmesi
3. Her öğrenci için sınıfta çalışmalarının toplanacağı ayrı bir kutu, dosya vb. oluşturulması
4. Her bir çalışmanın ve öğretmen kaydının üzerine tarih yazılması

SEÇME

1. Öğrenci seçimini kendi başına veya öğretmenin rehberliğinde yapabilir.
2. Portfolyonun türüne ve öğretmenin koyduğu koşullara bağlıdır.
   a. Süreci yansıtan portfolyolarda öğretmenin belirlediği konuları / gelişim alanlarını yansıtan çalışmaların seçilen örnekleri, öğrenme / gelişim sürecini yansıtacak şekilde yer alır.
   b. Ürünü yansıtan portfolyolarda öğretmenin belirlediği konular ile ilgili ortaya çıkan ürünler arasından seçilenler yer alır.

YANSITMA: Portfolyoyu herhangi bir çalışma dosyasından ayıran en önemli aşamadır. Bu aşamada öğrenci

1. Portfolyosuna seçtiği her bir çalışmayı niçin seçtiğini açıklar.
2. Çalışmalarını yaparken geçirdiği süreci ve bu süreçte öğrendiklerini anlatır.
3. Kendi başarısını görür, bunu ifade eder ve değerlendirme sürecine katılır.



Yansıtma Soruları:

- Bu çalışmayı nasıl yaptım?
- Bu çalışmadan ne öğrendim?
- Bu çalışmayı daha da geliştirebilir miyim? Nasıl?
- Çalışmalarım içinde en çok sevdiğim hangisi? Neden?
- Bana zor gelen bir çalışmam var mı? Varsa neden?
- Bu çalışmayı portfolyoma neden koydum?

SONUÇ

- Bu aşamada öğrenci "Bu çalışmayı niçin yaptık?" sorusunu yanıtlar.
- Okulda yaptığı çalışmalarla öğrendikleri arasında somut bağlar kurar.
- Tamamlanan portfolyo çalışmalarının öğrenci tarafından sınıf arkadaşları, öğretmeni ve ailesinden oluşan bir gruba sunumu yapılmalıdır. Portfolyonun sunumu, öğrencinin çalışmalarına önem vermesini sağlar ve kendine olan güvenini artırır.

3.3.2.2. PERFORMANS DEĞERLENDİRME:



Şekil 16: Performans değerlendirme araç ve yöntemleri

3.3.2.2.1. ÖZ DEĞERLENDİRME: Öğrencinin belirli bir konuda (*örneğin bir ürünü ortaya koymada gösterdiği performans vb.*) kendi kendisini değerlendirmesine denir.

- Öğrencilerin kendi özellikleriyle (yetenek, ilgi, beceri vb.) ilgili farkındalığının artmasını, zayıf ve güçlü yönlerini keşfetmesini sağlar.
- Öz düzenleme becerisi artar.
- Ölçütlü düşünme becerisi artar.
- Öğrenme motivasyonunu artırır.

Öz Değerlendirmenin Olası Dezavantajları: a. Öğrencilerin kendi performanslarını değerlendirirken yanlı davranmaları söz konusu olabilir. b. Başlangıçta deneyimsizlik nedeniyle performansın değerlendirilmesinde yanılgılar olabilir.

3.3.2.2.2. AKRAN DEĞERLENDİRME: Öğrencinin ortaya koyduğu performansa ilişkin arkadaşlarının değerlendirmesine denir.

- Akranların değerlendirme sürecine katılması nedeniyle daha katılımcı, aktif bir eğitim ortamı sağlanabilir (Sorumluluk duygusu artar.).
- Öğrencilerin eleştirel düşünme becerileri artar.
- Öğretmen dışındaki bir kaynaktan dönüt almak öğrencinin performansını artırabilir.
- Eleştiri kültürü (olumlu-olumsuz) gelişebilir.



Akran Değerlendirmenin Olası Dezavantajları: a. Öğrencilerin yanlı davranmaları söz konusu olabilir. Kişisel ilişkiler olumlu ya da olumsuz yönde değerlendirmeye etki edebilir. b. Genel izlenimle puan verme söz konusu olabilir.

3.3.2.2.3. RUBRİK (DERECELİ PUANLAMA ANAHTARI): Performans ürünlerinin değerlendirilmesinde en sık kullanılan araçlardan biridir. Amaç, öğretmen tarafından ürünün genel izlenimle puanlamasındaki öznelliğini azaltmaktır. Rubrikler ikiye ayrılır: a. Bütüncül (Holistik) rubrik b. Analitik rubrik

Tablo 5. *Bütüncül (Holistik) Rubrik Örneği (Maket Yapımı)*

Ölçüt	Puan
Maketi yaptığı materyal kalitelidir.	10
Ek yerlerinde pürüz bulunmamaktadır.	15
Özgün bir fikir ortaya koymuştur.	20
Maket çevre düzenlemesi gibi görsel öğelerle zenginleştirilmiştir.	15
Renklendirmeler uyumludur.	10
………	….
………	….
………	….
Toplam	100

Tablo 6. *Analitik rubrik örneği (Maket yapımı)*

Ölçüt	1 Puan	2 Puan	3 Puan	Puanı
Maketi yaptığı materyal kalitelidir.	Karton	Kontrplak	Masif	
Ek yerlerinde pürüz bulunmamaktadır.	>0.5 cm	0.3-0.5 cm	<0.3 cm	
Özgün bir fikir ortaya koymuştur.	Ders örnekleriyle aynı	Ders örneklerine yeni öğeler katmış	Tamamen özgün bir fikir	
Maket çevre düzenlemesi gibi görsel öğelerle zenginleştirilmiştir.	Zenginleştirme yetersiz	Ağaç, park vb. yapmış ancak yetersiz	Çevreyi tüm öğeleriyle yansıtmış	
Renklendirmeler uyumludur.	Renklendirmede özensiz	Kısmen uyumlu	Renklendirmeler çok iyi	
………	………	………	………	
………	………	………	………	
………	………	………	………	
Toplam				

- Rubrikler performans görevleriyle birlikte öğrenciye verilmelidir.
- Rubriklerin geliştirilmesi uzmanlık gerektirir. Rubrikler için geçerlilik ve güvenilirlik çalışmaları yapılmalıdır.

3.3.2.2.4. KONTROL LİSTELERİ: Gözlenen performans ürününün ölçütlere uygunluğunu “evet-hayır”, “var-yok”, “gösterdi-göstermedi” vb. bir biçimde kategorik (1-0) olarak puanlama amacıyla kullanılan araçlardır.

- Özellikle sergilenecek performans detaylı ve ardışık eylemler gerektirdiği zamanlarda kullanışlıdır. Örneğin bir deneyin eyleme dökülmesi vb.
- Pek çok işlem adımında oluşan performanstaki eksik adımları belirlemek için oldukça uygundur.

3.3.2.2.5. DERECELEME ÖLÇEKLERİ: Bu araçların kullanımında performansa dayalı işlemler ilk baştan sonuna kadar listelenir ve davranışın karşısına davranışın gösterilme derecesi en az üçlü [*örneğin tam gösterildi (3), kısmen gösterildi (2) ve gösterilmedi (1)*] bir biçimde derecelendirilir.



3.3.2.2.6. GÖZLEM FORMLARI: Öğrenme çıktılarının somut olarak gözlenebildiği bazı alanlarda bu yöntem oldukça kullanışlıdır.

- Özellikle fen derslerinde, meslek liselerinin somut performans ürünlerinin geliştirildiği vb. alanlar için oldukça uygundur.
- Gözlemler öğrenciler hakkında doğru ve hızlı bilgi elde edilmesini sağlar.
- Gözlem formları yarı yapılandırılmış biçimde olabileceği gibi tam yapılandırılmış bir formatta da olabilir.

3.3.3. GELENEKSEL VE DESTEKLEYİCİ YÖNTEMLERİN KARŞILAŞTIRILMASI

Tablo 7: Geleneksel ve destekleyici yöntemlerin karşılaştırmalı avantaj ve dezavantajları

GELENEKSEL YÖNTEMLER	DESTEKLEYİCİ YÖNTEMLER
Ürün değerlendirilir.	Süreç ve ürün birlikte değerlendirilir.
Öğrencinin ulaştığı noktanın tespiti önemlidir.	Ne öğrendikleri yanında, öğrendiklerini nasıl kullandıklarıyla ilgilenilir.
Essay dışında genellikle üst düzey düşünme becerilerinin değerlendirilmesinde yetersizdir.	Üst düzey bilişsel düşünme becerilerine odaklanır.
Değerlendirme öğrenmeden ayrıdır.	Değerlendirme öğrenmeyle bütünleşmiştir.
Bireyden ziyade gruba odaklıdır.	Odak noktası gruptan ziyade bireydir.
Başarının bireyler arası değerlendirilmesine odaklıdır.	Öğrencinin bireysel olarak gelişimine odaklıdır.
Geçerlilik ve güvenilirlik kontrolü daha kolaydır.	Geçerlilik ve güvenilirlik problemi olabilir.

4.TEST VE MADDE İSTATİSTİKLERİNE GENEL BİR BAKIŞ

Doç. Dr. Güçlü ŞEKERCİOĞLU

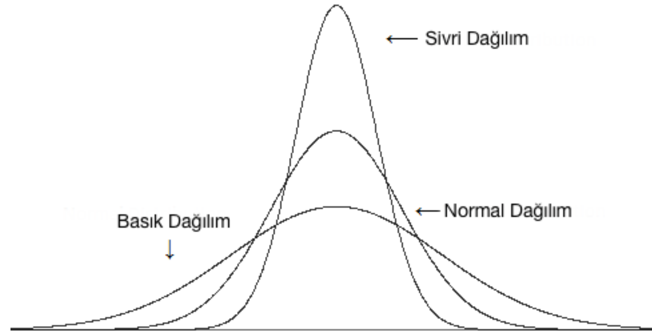
4.1. TEST İSTATİSTİKLERİ: Ölçme işlemi sonucunda bireylerin toplam puanları üzerinden hesaplanan istatistiklerdir. Eğitimde en sık kullanılan istatistikler merkezî eğilim ölçüleri, değişkenlik ölçüleri, dağılım özellikleri, standart puanlar vb.dir.

Merkezî eğilim ölçüleri: aritmetik ortalama, medyan (ortanca) ve moddur. Bu üç istatistiğin de temel işlevi bir puan dizisindeki merkezi bulmaktır. Merkezî eğilim ölçüsü hesaplamak ne işimize yarar? Ne tür kararlar almamızı sağlar? Merkezî eğilim ölçüleri, örneğin ortalama, grubun başarısı (ya da hangi özellik gözleniyorsa) hakkında kabaca da olsa fikir yürütmemizi sağlar. Söz gelimi ortalama 53 olursa grup başarı düzeyi açısından vasat, 82 olursa başarılı, 27 olursa başarı düşük vb. değerlendirmeler yapılabilir. Üstelik bazen bu ölçülerden herhangi biri ölçüt olarak kabul edilerek değerlendirme (bağıl) yapılabilir. Ancak bu istatistikler ile grup hakkında sınırlı değerlendirmeler yapılabilir.

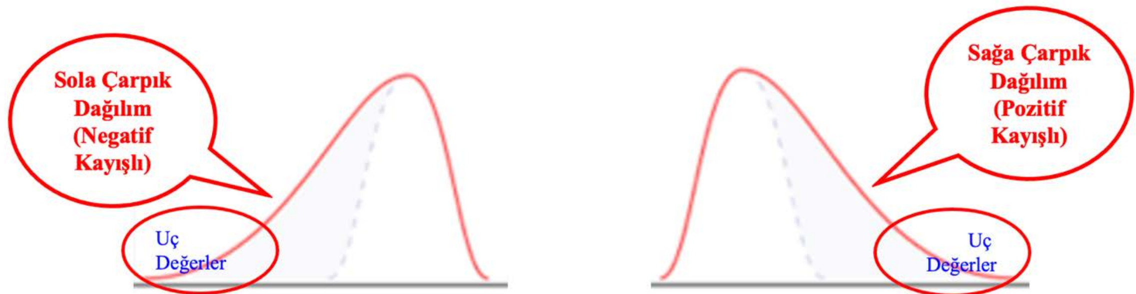
Diğer taraftan değişkenlik ölçüleri ranj (dizi genişliği), varyans, standart sapma ve çeyrek sapmadır. Değişkenlik ölçüleri grubun kabaca homojen ya da heterojen bir özellik gösterip göstermediği, standart sapma örnek olarak verilirse değişkenliğin ortalama etrafında nasıl dağıldığını gösterir. Puanların değişkenliği, grup hakkında biraz daha ayrıntılı bir değerlendirme yapılmasını sağlar.



Puanların dağılım özellikleri yine grup hakkında bilgi elde etmemizi sağlayan istatistiklerdir. Bu istatistikler genellikle histogramlar ya da çizgi grafikleri ile gösterilse de grubun dağılımı hakkında değerlendirme yapmamıza olanak sağlayan pek çok istatistik bulunmaktadır.



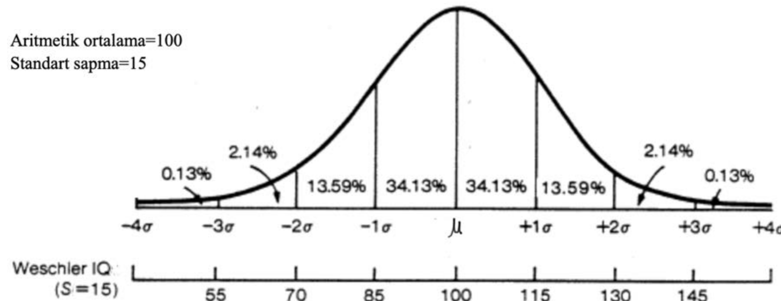
Şekil 17: Normal, sivri ve basık dağılımlar



Şekil 18: Sola ve sağa çarpık dağılımlar

Evrende pek çok fiziksel, fizyolojik ya da psikolojik özelliğin normal dağılım eğrisine göre dağıldığı, diğer bir ifadeyle bu dağılımın "*normalin bir ölçüsü*" olduğu kabulü yapılır. Sivri, basık, sola çarpık ve sağa çarpık dağılımlar normalden sapma gösteren dağılımlardır. Normalden sapma durumlarının pek çok nedeni olabilir. Örneğin grup sayısının az olması, testin kolay ya da zor olması, grubun ölçülen özellik açısından genel niteliği vb. durumlar.

Normal dağılım ortalama ve standart sapma gibi basit iki istatistikle özetlenebilir. Normal dağılım ortalama etrafında üç standart sapma sağa ve sola, toplamda altı alanda özetlenebilir (%99,74'lük kısmı). Dolayısıyla bir kişinin ortalamadan kaç standart sapma uzaklaştığı, grup içerisindeki yerini gösterir.

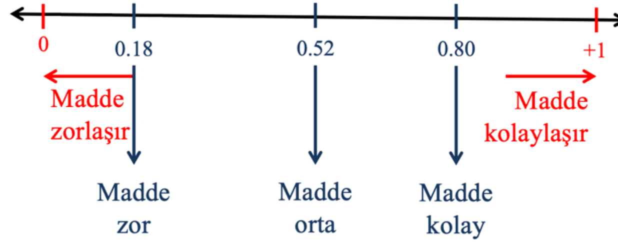


Şekil 19: Standart normal dağılım altında yer alan alanlar



4.2. MADDE İSTATİSTİKLERİ: Ölçme işlemi sonucunda bireylerin madde puanları üzerinden hesaplanan istatistiklerdir. Eğitimde en sık kullanılan istatistikler madde güçlük indeksi, madde ayırt edicilik indeksi, madde güvenilirliği vb.dir.

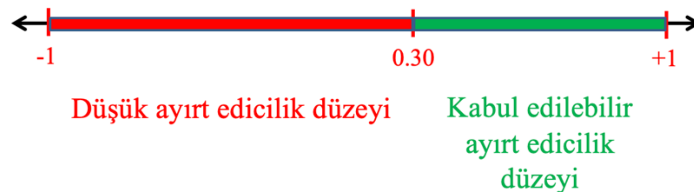
Madde güçlük indeksi; maddenin kabaca zor mu, orta mı, kolay mı olduğunu gösteren bir istatistiktir.



Şekil 20: Madde güçlük indeksi

Madde güçlük indeksi 1'e yaklaştıkça maddenin kolaylaştığı, 0'a yaklaştıkça zorlaştığı anlaşılır. 0.50 civarı ise ortalama güçlükte olduğu biçiminde yorumlanır. Her ne kadar ayrı bir formülü olsa da madde güçlükleri toplamının madde sayısına bölümü testin ortalama güçlüğünü ortaya koyar. Bu noktada öğretmenlerin zihninde şu sorular olabilir: "*Hazırbulunuşluluğu ölçmeyi ya da öğrenme eksiklerini tespit etmeyi ya da not vermeyi amaçlayan bir öğretmen testinin / sınavının ortalama güçlüğünü ne düzeyde tutmalıdır? Soruların güçlük dağılımı nasıl olmalıdır?*" Bu tür amaçlarla hazırlanan sınavlarda kolay ve zor sorular teste konulabilir ancak ortalama güçlükteki sorular her zaman sayıca daha fazla olmalıdır. Nihayetinde testin ortalama güçlüğü hesaplandığında 0.50 olması ideal durumdur. Pekâlâ "*test zor ya da kolay olursa ne olur?*" Bir test zorlaştıkça dağılımın sağa çarpık, kolaylaştıkça sola çarpık olma olasılığı artar. Diğer bir ifadeyle grup başarı düzeyi açısından homojenleşmeye başlar ve homojen bir dağılımda testin bilenle bilmeyeni birbirinden ayırt etme gücü azalabilir ve güvenilirlik katsayısı düşebilir. Diğer bir ifadeyle ayırt ediciliği ve güvenilirlik katsayısı yüksek bir ölçme işlemi için öğretmen zor ve kolay sınavlardan ziyade orta güçlükte bir test hazırlamalıdır (Daha önce de açıklandığı gibi elbette testte kolay ve zor sorular olabilir.).

Madde ayırt edicilik indeksi ise maddenin bilenle bilmeyeni ayırt edip edemediğine ilişkin bir istatistiktir. -1 ile 1 arasında değer alır ve genellikle alanyazında 0.30 ile 1 arası ayırt edici olarak nitelendirilir.



Şekil 21: Madde ayırt edicilik indeksi



1. EĞİTİM İZLEME ARAŞTIRMALARI: TARİHİ, ÖNEMİ VE TÜRKİYE'DEN SONUÇLAR

Dr. Hayri Eren SUNA

Bu ders kapsamında, günümüzde önemi gittikçe artan ve daha sık kullanılan uluslararası eğitim izleme ve çalışmalarıyla ilgili detaylı bilgiler sunulacaktır. Öncelikle 1960'lı yıllardan bu yana uygulanan ve uluslararası düzeyde gittikçe daha fazla önem atfedilen izleme araştırmalarının tarihi hakkında ardından da Türkiye'nin katıldığı çalışmalar hakkında detaylı bilgiler verilecektir. Bu kapsamda izleme araştırmalarıyla ilgili cevaplamaya çalışacağımız sorular: Nasıl geliştiriyorlar? Nasıl uygulanıyor? Neden bu kadar popüler hâle geldi? Türkiye'nin zaman içerisindeki söz konusu çalışmalardaki gelişimi nasıl gerçekleşti? Performansı nasıl değişti? Millî Eğitim Bakanlığı, eğitim sisteminin performansını izlemek için farklı paydaşlarla nasıl çalışmalar yürütüyor? Sorularına birlikte cevap vereceğiz.

Kavramlar

İlk kısımda genel olarak kavramların nasıl ele alındığı ve bu kavramlar hakkındaki farklı görüşleri ele alarak başlamak faydalı olacaktır. Bu kavramlardan ilki, sıklıkla başarı (achievement) ya da bazı kaynaklarda erişi olarak ifade edilmektedir. Başarı kavramı aslında günlük hayatımızda da çok yoğun kullandığımız, farklı anlamlar atfettiğimiz ve genellikle olumlu bir özellik olarak değerlendirilmektedir. Aslında eğitim alanında biz başarıyı, belirli bir konu alanında veya uygulama alanında kazandırılan/kazandırılmaya çalışılan kazanımlara bireylerin ulaşma/erişme düzeyleri olarak anlamlandırıyoruz. Bu nedenle başarı aynı zamanda erişi olarak tanımlanıyor. Başarı, konu alanında verilen eğitim sonucunda öğrencilerin beklenen kazanımlara, becerilere ya davranışlara sahip olma düzeyini ifade eden bir kavram. Özetlemek gerekirse erişi ya da başarı kavramı, sunulan bilgilere, sunulan içeriğe o eğitimi alan bireylerin ne kadar ulaşabildiğine dair bir gösterge, erişim düzeyi olarak ifade edilmektedir.

Bir diğer kavram olan beceri ise başarıdan farklı olarak tekil kazanımlardaki erişim düzeylerinden ziyade öğrencilerin bilgi ve becerilere sahip olduktan sonra bunları ne kadar kullanabildiğine dair bir gösterge olarak değerlendirilmektedir. Beceri; bilgilerin bir araya toplanıp, doğru bilgilerin seçilip ardından seçilen bilgilerin doğru şekilde kullanılması, gerçek ya da tasarlanan bir durumda kullanılabilmeyetisini ifade etmektedir.

Eğitimin daha geleneksel olduğu zamanlarda bilgiye sahip olmak, bilgiyi ezberde tutmak ya da yeri gelince hatırlamak ve anlamak kıymetliydi. Bilgiye ulaşmanın gittikçe kolaylaştığı günümüzde artık bu bilgileri uygun durumlarda kullanmak önem kazanmakta ve beceri kavramının önemi öne çıkmaktadır.

Değineceğimiz son kavram ise okuryazarlık (literacy). Okuryazarlık da aslında beceriyle çok ilişkilendirilen ve birçok kaynakta bir beceri türü olarak ifade edilen bir yapıdır. Günümüzde dünyada önem atfedilen eğitim izleme araştırmalarından birisi olan Uluslararası Öğrenci Değerlendirme Programında (PISA) ölçülen özellikler "okuryazarlık" olarak tanımlanmaktadır. Okuryazarlık, özellikle belirli bir alanda ilgili doğru bilgileri bulma, bu bilgileri yanlış bilgilerden ayıklama, bunun için temel okuryazarlık becerilerini kullanma, ardından doğru bilgileri seçerek ve uygun analiz yöntemlerini kullanarak gerçek ya da tasarlanmış durumlarda kullanma becerisi olarak tanımlanmaktadır. Eğitim sistemimiz de dâhil olmak üzere birçok çağdaş eğitim sisteminde, aslında öğrencilere salt bilginin yanında çeşitli beceriler kazandırmak ve kazandırılan bu becerileri doğru ortamlarda, doğru yorumlayacak ve kullanabilecek şekilde gerekli tüm altyapıyı kazandırmak amaçlanmaktadır.

Bilimsel ve teknolojik gelişmeler sonucunda bilgiye ulaşmanın kolaylaşması, 2000'li yıllardan sonra eğitim sistemlerinin öğrencilere bilgiye ulaşmanın ötesinde bunları uygulama becerisi kazandıracak şekilde revizyona gitmesini gerektirmiştir. Birçok eğitim sistemi artık daha fazla ve daha ağırlıklı şekilde öğrencilerine gerçek hayatta karşılaşacakları problemleri



çözebilecek becerileri kazandırmaya çalışmaktadır. Bu beceriler yalnız eğitim açısından değil, eğitimin ötesinde bireylerin artık günümüzde istihdam edilebilirliklerini artırmak için de son derece önemlidir. Eğitimin yanında iş gücü piyasasında da bilgiye ulaşmanın yanında mevcut problemleri çözebilen, bunun için doğru kişilerle iletişim kurabilen, iş birliği yapabilen ve bu süreçte kendi akademik ve sosyal becerilerini kullanabilen insanlara ihtiyaç duyulmaktadır.

Güncel eğitim politikaları da talep edilen bu becerilere cevap vermek adına eğitim programlarını ve sundukları tüm eğitim yaşantılarını buna göre tasarlamak durumundadır. İş gücü piyasasında artık otomasyon yapay zekâ çalışmaları hızla yayılmakta ve her geçen gün daha fazla sayıda meslek, daha fazla sayıda iş otomasyon tarafından -yani makineler tarafından- yapılır hâle gelmektedir. Hâliyle böyle bir dünyada iş gücü piyasasında rutin işlerden ziyade daha üst düzey bilişsel becerilere sahip, iş birliğine açık, yeniliklere daha kolay adapte olabilen ve mevcut problemleri çözme konusunda yüksek yetkinliğe sahip bireylerin istihdam edilebilirliği artmıştır.

Yine hatırlamamız gerekiyor ki bu becerilere sahip insan kaynağının yetişmesi, ülkelerin en büyük önceliğidir. Eğitim de öğrencilere söz konusu becerilerin kazandırılmasında ülkelerin sahip olduğu en önemli araçtır. Bu nedenle ülkeler eğitim sistemlerine daha fazla daha yatırım yaparak ve gerekli iyileştirmeleri sağlayarak öğrencilerine gerekli becerileri kazandırmaya, onların okuryazarlıklarını geliştirmeye gayret etmektedir.

Uluslararası İzleme Çalışmaları

Uluslararası izleme çalışmaları ne zaman, ne amaçla ortaya çıktı? Eğitim izleme çalışmaları, bugün çok popüler olmakla beraber aslında çok uzun süredir sahada olan uygulamalardır. Bugün birçok ulusal izleme araştırmasına temel teşkil eden ilk çalışamalardan biri 1964 yılında gerçekleştirilen uluslararası matematik izleme çalışmasıdır. TIMSS’in de geliştiricisi olan kurum (IEA) tarafından gerçekleştirilen bu çalışma 1960’larda matematik üzerinde başlamıştır.

Aslında bu tarz izleme çalışmalarına ihtiyaç duyulmasının ve bu çalışmaların popüler olmasının sebeplerinden ilki karşılaşıtırılabilir veri ihtiyacıdır. Bildiğiniz üzere her ülke eğitime önem vermekte ve eğitimi önemli bir yatırım alanı olarak görmektedir. Bununla birlikte ülkelerin kendi insan kaynağını geliştirmede attığı adımlar birinden diğerine önemli farklılıklar gösteriyor. Bu farklılıklar elbette kaçınılmazdır zira ülkelerin kültürel özellikleri, eğitime yaklaşımları ve eğitimden beklentileri kısmen farklıdır. Diğer yandan da farklı yaklaşımlar izleyen ülkeler şu geri bildirimlere ihtiyaç duyuyor: Biz bu yaptıklarımızla eğitimi nereye götürüyoruz? Diğer ülkelerden ne gibi farklılıklara sahibiz? Yaptıklarımızla eğitimde istediğimiz ivmeyi yakalayabildik mi, yakalayamadık mı? Bu soruları yanıtlamak için ise ülkeler arasında karşılaştırılabilir veriye ihtiyaç duyuluyor. Eğitim izleme araştırmaları da 1960’lı yıllardan itibaren özellikle bu ihtiyacı karşılıyor.

1960'lı yıllarda matematik alanı ile başlayan çalışmaların bugün kapsamı büyük ölçüde genişlemiştir. Örneğin PISA ile beraber değerlendirilen alanlar artmış; matematik, fen ve okuma becerileri üç ana alan olarak belirlenmiştir. Uluslararası Matematik ve Fen Eğilimleri Araştırmasında (TIMSS) matematik ve fen, Uluslararası Okuma Becerilerinde Gelişim Araştırmasında (PIRLS) ise okuma becerileri değerlendirilmektedir. Ayrıca izleme çalışmalarının kapsamı günümüzde sosyal ve duygusal beceriler, vatandaşlık becerileri, yetişkinlik becerileri, dijital okuryazarlık gibi çok farklı alanları da içerecek şekilde genişlemiştir. Ülkelerin eğitim performanslarına dair artan geri bildirim ihtiyacını karşılamak için daha fazla izleme araştırması yapılıyor. Bu geri bildirimler sadece miktar olarak daha fazla değil, artık daha fazla alanda ve daha fazla faktöre dair geri bildirim sağlanmaktadır.

İzleme araştırmalarının sağladığı ikinci katkı, ihtiyaç duyulan bağlamsal ve bütüncül ilişkilere dair bulgular sunmasıdır. Çalışmalarda sadece eğitim performansının nasıl değiştinin ötesinde elde ettiğimiz bilginin kapsamı ve bağlamı bundan çok daha fazlasıdır. Aslında bu çalışmalar öğrenci performansındaki değişimin birçok özellikle ilişkili olup olmadığı konusunda



da bilgi vermektedir. Örneğin öğrencilerin akademik başarıları okul iklimi, eğitim yatırımları, öğretmen özellikleri, öğrencilerin sosyoekonomik özelliklerine göre nasıl değişiyor gibi önemli sonuçları yine bu çalışmalar sayesinde görebiliyoruz. Üstelik bu bilgiler yalnız Türkiye'ye ilişkin değil, farklı ülkelerdeki durumu da değerlendirebileceğimiz şekilde veriliyor. Dolayısıyla veriler eğitim politikacıları, eğitim araştırmacıları ve eğitime ilgi duyan tüm paydaşlar açısından son derece önemlidir. Ayrıca bu çalışmaların hepsinin verileri de açık hâlde bulunuyor, sizler de bu verileri temin edip kendi araştırmalarınızı yapabilirsiniz.

Son olarak izleme çalışmaları, eğitim politikalarının bir bütün olarak oluşturduğu etki konusunda da önemli bulgular sağlamaktadır. Bu çalışmaların hiçbiri doğrudan nedensel çıkarımlar yapmaya uygun çalışmalar değildir. Yani bir eğitim sisteminde yapılan işlemler ve bu işlemlerin ayrı ayrı etkilerine dair sonuçlar sağlamamaktadır. Ancak belirli zaman içinde -genelde üç ve dört yıllık periyotlarla- eğitim sisteminin farklı bileşenlerine dair önemli çıktılar sağlamaktadır. Dolayısıyla çalışmalar, eğitim performansının zaman içinde nereden nereye geldiği hakkında bütüncül bir değerlendirmeye imkân sağlamaktadır.

Günümüzde En Yoğun Katılım Gösterilen Çalışmalar

PISA, Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) tarafından geliştirilen ve üç yıllık periyotlarla uygulanan izleme çalışmasıdır. OECD, Türkiye'nin kurucu üyesi olduğu bir örgüt olup temel amacı tüm üye ülkeler arasındaki ekonomik iş birliklerini ve kalkınmayı destekleyecek programlar oluşturarak birlikte kalkınmayı sağlamaktır. Bu bağlamda OECD, eğitimi, ekonomik kalkınmaya katkı sunan ve çağdaş vatandaşlık becerilerine sahip insan kaynağının geliştirilmesinde en önemli araç olarak görmektedir.

Bir diğer izleme çalışması olan TIMSS ise Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA) tarafından geliştirilmekte ve dört yıllık periyotlarla uygulanmaktadır. Türkiye uzun süredir hem PISA'ya hem de TIMSS'e katılım göstermektedir. Son olarak PIRLS de yine IEA tarafından beş yıllık periyotlarla gerçekleştirilen bir izleme çalışmasıdır. Türkiye uzun bir aradan sonra 2021'de tekrar PIRLS'e katılmış ancak bu uygulamanın sonuçları 2022 yılının sonunda açıklanacaktır.

2. PISA KAPSAMI VE SONUÇLARI

Dr. Hayri Eren SUNA

Bu ders kapsamında PISA'nın nerede, nasıl uygulandığı ve hangi bağlamlarda değerlendirme yaptığı açıklandıktan sonra Türkiye'nin performansı tartışılacaktır.

OECD, ekonomik kalkınmanın ve sürdürülebilirliğinin sağlanmasındaki en önemli kaynak olan beşerî sermayeyi üreten faktör olarak eğitimi konumlandırmaktadır. Eğitimin ekonomik kalkınma üzerindeki belirleyici rolü dolayısıyla eğitim niteliğini detaylı olarak ele alan ve üç yıllık periyotlarla uygulanan PISA çalışmasını gerçekleştirmektedir. En güncel PISA uygulaması olan 2018 döngüsüne 79 ülke/ekonomi katılım göstermiştir. Dünyanın dört bir yanından farklı coğrafyalardan, farklı kültürlerden, farklı dinamiklere sahip eğitim sistemleri olan ülkeler yaygın bir şekilde PISA'ya katılım göstermektedir.

PISA, 15 yaş grubunu hedef grup olarak almaktadır. Bunun nedeni birçok Avrupa ve OECD ülkesinde 15 yaşa tekabül eden dönemin yaklaşık olarak ortaokulun sonuna denk gelmesi ve bu dönemde zorunlu eğitimin tamamlanmasıdır. Dolayısıyla PISA, zorunlu eğitim dönemini bitiren öğrencilerin henüz iş gücü piyasasına katılmadan ya da eğitimlerine devam etmeden matematik, fen ve okuma becerileri alanlarında hangi yeterliklere sahip olduğuna dair bilgi vermektedir. Yüksek katılıma bağlı olarak 79 ülkeden 600 binden fazla öğrenciye dair verilerin kullanıldığı PSA 2018, ülkelere dair karşılaştırmalı çıkarımlar sağlamaktadır.

PISA, 2000 yılından bu yana uygulanmakta ve her üç yıllık döngüde bir alana ağırlık verilmektedir. 2000 yılında ağırlıklı alan okuma becerileri iken 2003 yılında matematik



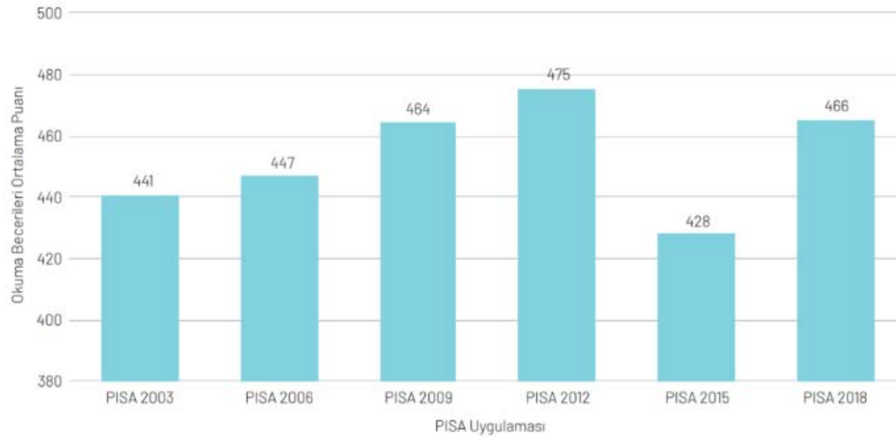
okuryazarlığı ve 2006 yılında fen okuryazarlığı, 2018 yılında tekrar okuma becerileri olacak şekilde çapraz bir modelleme kullanılmaktadır. Ağırlıklı alan olarak belirlenen alanda değerlendirme çerçevesi ve tüm değerlendirme süreci yenilenmekte ve derinlemesine analizler yapılmaktadır. Her döngüde bir alana ağırlık verilmekle beraber matematik okuryazarlığı, fen okuryazarlığı ve okuma becerileri alanlarının hepsine dair ölçme yapılmaktadır. OECD, ekonomik ve kalkınma alanındaki dinamiklerin günümüzde sıkça değişmesi nedeniyle eğitim sistemlerine ve ülkelere mümkün olduğunca daha sık geri bildirim vermek için PISA uygulamalarının üç yılda bir yürütüldüğünü ifade etmektedir.

PISA 2018 uygulamsına ait Türkiye örneklemi incelendiğinde, 15 yaş grubunda kısmen 8. sınıfta bulunan ortaokul son sınıf öğrencilerimiz bulunmakta ama yoğun olarak 9 ve 10. sınıf öğrencilerimizin olduğu görülmektedir. Okul türlerine göre dağılım incelendiğinde ise öğrencilerimizin %44'ünün Anadolu lisesi, %31'inin mesleki ve teknik Anadolu lisesi ve %14'ünün Anadolu imam hatip lisesinde eğitim gördüğü, bu dağılımın ülkemizde hâlihazırda lise türlerindeki öğrenci dağılımıyla benzerlik gösterdiği ifade edilebilir.

Örneklem seçimi yedi coğrafi bölgeye göre değil, Türkiye'nin sosyoekonomik farklılıklarını da dikkate alan 12 bölgeden oluşan İBBS Düzey-1 istatistiksel örnekleme sistemine göre yapılmaktadır. Bu örnekleme sürecinde bölgesel öğrenci yoğunlukları dikkate alınmakta ve daha büyük illerimizin olduğu bölgeler daha fazla öğrenciyle, daha az olduğu yerler ise daha az öğrenciyle temsil edilmektedir.

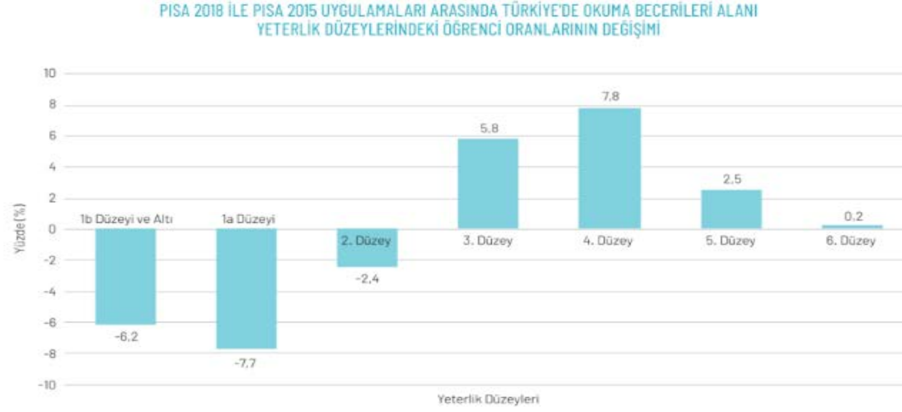
Ortalama Puanlardaki Değişim

PISA 2003 İLE PISA 2018 UYGULAMALARI ARASINDA TÜRKİYE'NİN OKUMA BECERİLERİ PERFORMANSINDAKİ DEĞİŞİM





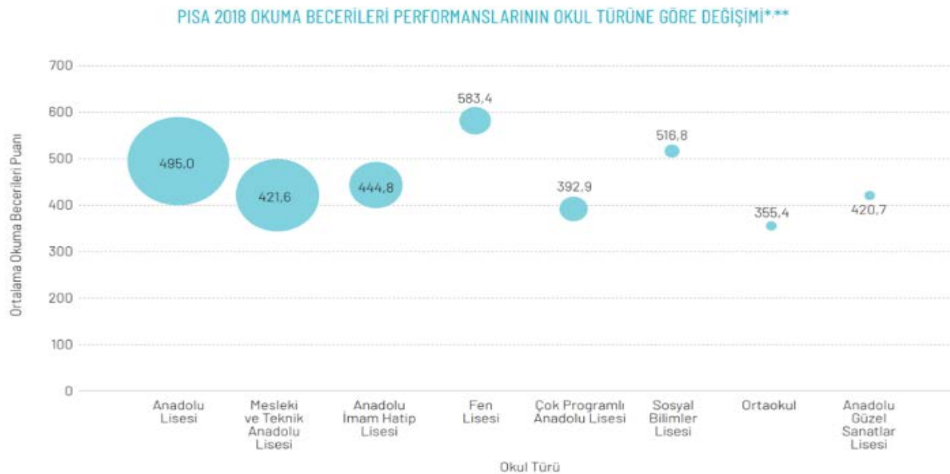
Yeterlik Düzeylerindeki Değişim



Şekil 1. Okuma becerisi ortalama puanlarındaki ve yeterlik düzeylerindeki değişim

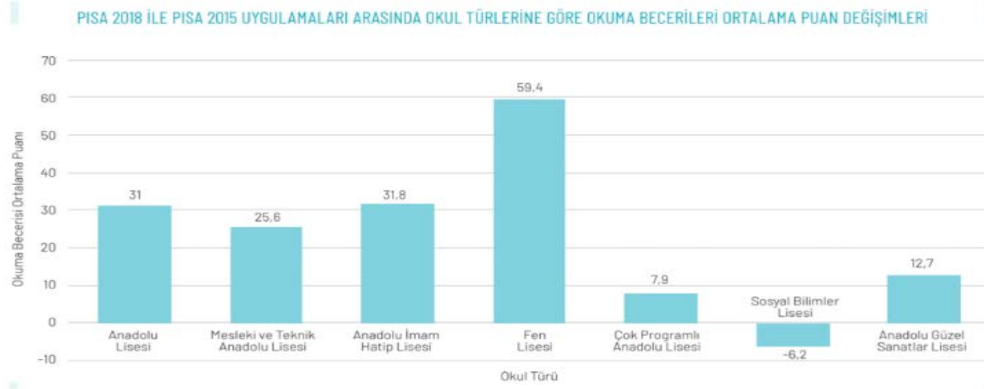
Şekil 1'de görüldüğü gibi Türkiye, PISA 2003 yılından bu yana katılım göstermekte olup 2003 senesinde 441 olan okuma puanı genel bir artış eğilimi göstererek 466 puana ulaşmıştır. 2003'ten 2012'ye kadar kesintisiz bir artış, 2015 döngüsünde bir düşüş ve söz konusu 2018 döngüsünde tekrar artış vardır. PISA'nın en önemli çıktılarından bir tanesi öğrencilerin hangi yeterlik düzeyinde olduğuna dair veriler sunmasıdır. Yeterlik düzeyleri son derece önemlidir çünkü her bir yeterlik üzerinde bulunan öğrencinin o alanda ne yapıp ne yapamadığını, davranış bazında hangi davranışları gösterip hangi davranışları gösteremediğine dair somut açıklamalar yapmaktadır. Örneğin en üst düzeye yeterlik düzeyi olan 6. düzeydeki bir öğrencinin çok karmaşık metinleri rahatlıkla okuyabilmekte, anlamlandırabilmekte, metin içerisindeki dinamiklerin farkına varıp bunları ilişkilendirebilmekte, farklı metinlerdeki örtülü anlamları fark edebilmektedir. 2015'ten 2018'e puan seviyesindeki artışın en memnun edici yanı, en düşük yeterlik düzeyleri olan 1-b, 1-a ile kısmen 2. yeterlik düzeylerindeki öğrenci oranları azalırken daha üst yeterlik düzeylerindeki öğrenci oranlarımızın artmasıdır.

Okul Türlerine Göre Puan Ortalamaları





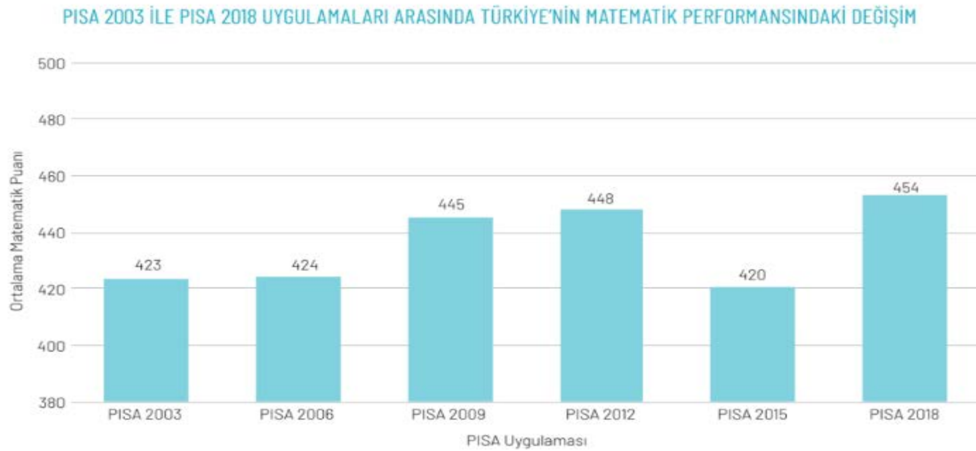
Önceki Döngü ile Performans Farkı



Şekil 2. Okul türlerine göre okuma becerisi puan ortalamaları ve performans farkları

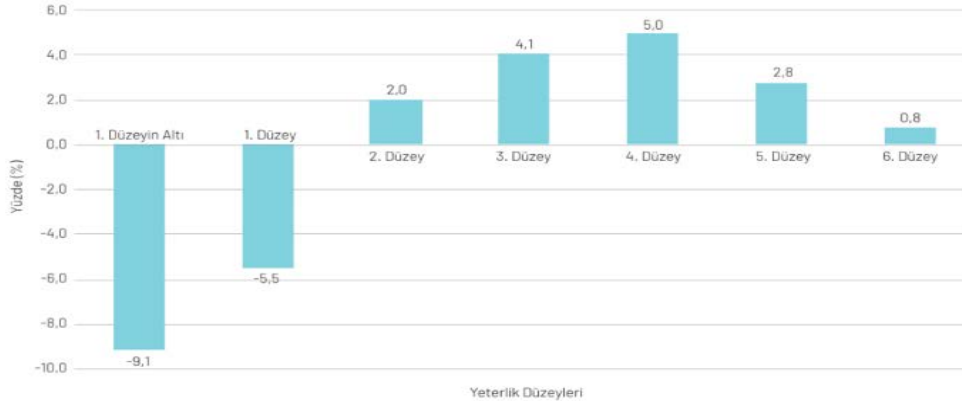
Okul türlerinin okuma becerisi puan ortalamalarının nasıl farklılıklar gösterdiği incelendiğinde, Şekil 2'de görüldüğü gibi en başarılı okul türünün olan fen lisesi 583 puan ile OECD genel ortalamasının da oldukça üzerindedir. Şekil 2'nin sol tarafında bulunan dairelerin büyüklüğü öğrenci sayısının büyüklüğünü temsil etmektedir. O yüzden en büyük daireler Anadolu Lisesi, mesleki ve teknik Anadolu ve Anadolu imam hatip lisesine aittir. Performans açısından fen liselerini sosyal bilimler liselerinin takip ettiği gözlenmiştir. Okullarımız arasında başarı farkları bulunmakla beraber Bakanlığımız okullar arasındaki başarı farklarını en aza indirmek adına önemli adımlar atmaktadır. Burada bir diğer önemli gösterge de geçen yıllara kıyasla okullar arasındaki başarı farkının kısmen azalmış olmasıdır. Elbette bu farkın daha da azalması hedeflenmektedir. 2015'ten 2018'e performans artışı ve yeterlik düzeylerindeki iyileşme sosyal bilimler liseleri dışında tüm okul türlerimizde gözlenmiştir. Örneğin fen lisesiöğrencilerinin ortalaması 583 puan olup 2015 PISA uygulamasından 59 puan daha yüksektir.

Ortalama Puanlardaki Değişim





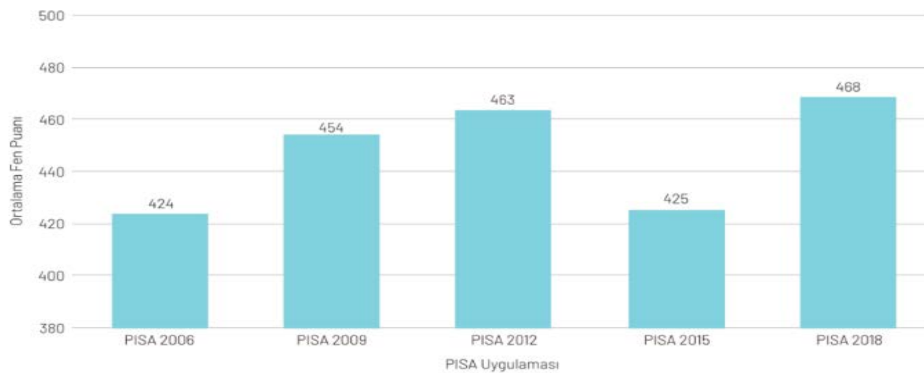
Yeterlik Düzeylerindeki Değişim



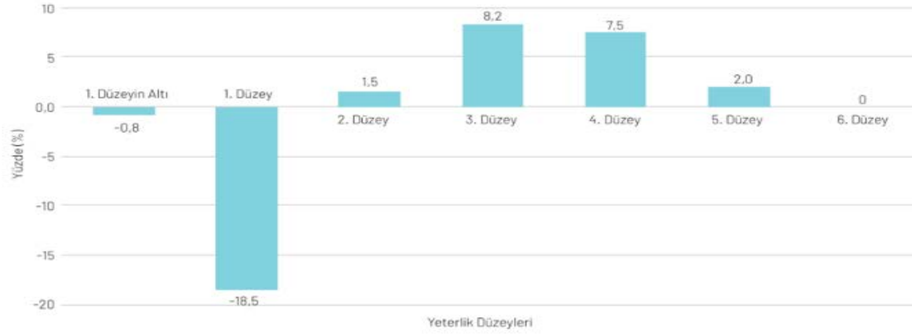
Şekil 3. Matematik okuryazarlığı ortalama puanlarındaki ve yeterlik düzeylerindeki değişim

Matematik okuryazarlığına geldiğimizde, Şekil 3'te görüldüğü gibi sürece 2003 yılında 423 puanla başlayan Türkiye'nin ortalama matematik okuryazarlık puanının on beş yıllık aradan sonra 454 puana çıktığı görülmektedir. Okuma becerisinde olduğu gibi matematik okuryazarlığında da OECD ortalamasının kısmen altında bulunmakla birlikte 1. düzey ve bu düzeyin altındaki öğrenci oranları azalırken daha yüksek performansı ifade eden 3, 4 ve 5. düzeydeki öğrenci oranlarında artış gözlenmiştir. Okuma becerisine benzer olarak matematik okuryazarlığı başarısında da fen liselerini; sosyal bilimler, Anadolu, mesleki ve teknik Anadolu ve Anadolu imam hatip liseleritakip etmektedir. Fen liseleri matematikte OECD ortalamasından çok daha yüksek performans göstermektedir. Okullarımız arasında başarı farkını en aza indirmek için diğer projelerle birlikte matematik seferberliği gibi somut adımlar atılmaktadır. Son olarak çok programlı Anadolu Lisesi ve sosyal bilimler liselerinde kısmi puan düşüşü bulunmakla birlikte diğer tüm okul türlerinde önemli puan artışları gözlenmiştir.

Ortalama Puanlardaki Değişim







Şekil 4. Fen okuryazarlığı ortalama puanlarındaki ve yeterlik düzeylerindeki değişim

Son olarak fen okuryazarlığı alanında 2006'da Türkiye 424 puanla başladığı süreçteki performansını 2012'ye kadar önemli ölçüde artırmış, 2018'de de artırmaya devam ederek en üst düzeye ulaştırmıştır. Birinci düzey ve altında bulunan yaklaşık %20'lik öğrenci artık daha yüksek düzeylerde yer almaktadır ve bu durum 43 puanlık artışın sonucudur. En başarılı iki okul türü olan fen ve sosyal bilimler liselerini; Anadolu, mesleki ve teknik Anadolu ile Anadolu imam hatip liseleri takip etmektedir. Yine 2018 ile beraber sosyal bilimler haricindeki tüm liselerde önemli puan artışları gözlenmiştir.

Özetle Türkiye, matematik ve fen okuryazarlığında son 15 yılda, katıldığı tüm PISA uygulamalarından en yüksek ortalama puana ulaşmış; okuma becerilerinde ise performansını önemli ölçüde artırmıştır. Elbette hedef OECD ortalamasına ulaşmak ve OECD ortalamasını geçmektir ama öğrencilerimizin puanlarını artırmış olması olumlu bir gelişmedir.

3. TIMMS KAPSAMI VE SONUÇLARI

Dr. Hayri Eren SUNA

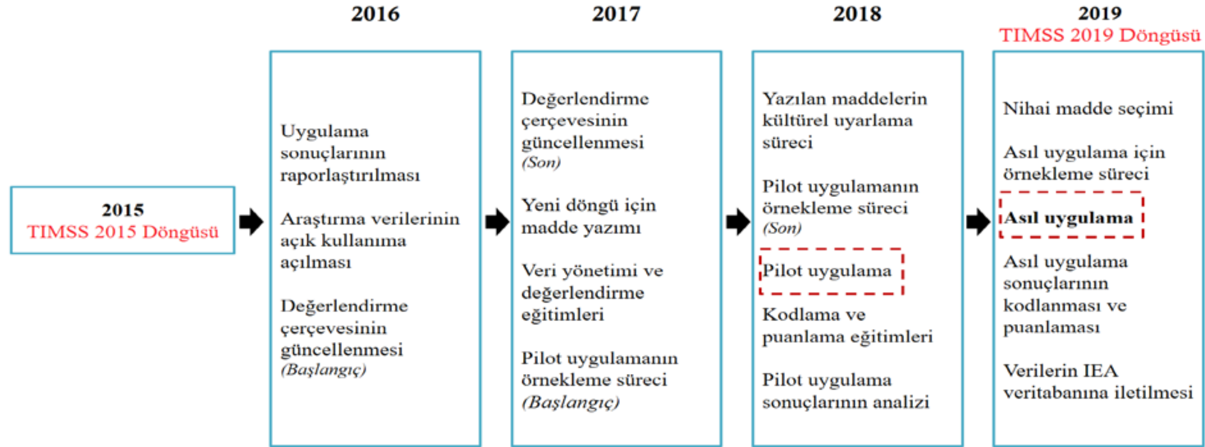
Bu ders kapsamında TIMSS izleme araştırmasının tasarımı, uygulama şekli ve bulguları üzerinde durulacaktır. TIMSS, IEA tarafından dörder yıllık döngüler şeklinde gerçekleştirilmektedir. PISA ile TIMSS arasındaki en temel fark, PISA'da zorunlu eğitim dönemini tamamlayan öğrencilerin iş gücü piyasası tarafından ve ekonomik kalkınma açısından istenen, uygulamaya dönük okuryazarlık becerilerinin hangi seviyede olduğu değerlendirilmekte iken TIMSS'te eğitim programı (curriculum based) ile ilişkili becerilerin değerlendirilmesidir. TIMSS'te 4 ve 8. sınıf düzeyindeki öğrencilerin ilgili dönemin eğitim programlarından beklenen özelliklere hangi düzeyde sahip olduklarına dair matematik ve fen alanlarında ayrı ayrı uygulanmalar yapılmakta ve ilgili eğitim programlarında öğrencilerin okul başarısına dair daha doğrudan çıktılar elde edilmektedir.

TIMSS, ilk defa 1995 yılında uygulanmaya başlanmış olup 2019'da araştırmanın yedinci döngüsü tamamlanmıştır. Türkiye TIMMS'e 1999 yılında ilk defa 8. sınıf düzeyinde katılım gösterirken 2011 yılından bu yana son üç döngüye hem 8. sınıf hem de 4. sınıf düzeyinde katılmıştır. TIMSS'in son döngüsü olan 2019 döngüsünde tüm dünyadan 580.000 öğrenci, 310.000 veli, 19.000 yönetici ve 52.000 öğretmenden veri toplanmıştır. Araştırmanın yedinci



döngüsü toplam 64 katılımcı ülkeyle gerçekleştirilmiş ve PISA çalışmasıyla son derece yakın istatistikler elde edilmiştir.

Biraz da uygulama süreci ve hazırlıklardan bahsetmek gerekirse TIMSS 2019 uygulaması için yapılan hazırlıklara 2016 yılında başlanmaktadır. Şekil 5'te bir TIMSS döngüsünün uygulama öncesi ve sonrasında yapılan işlemler ve hazırlıklar yıl bazında verilmiştir.



Şekil 5. Bir TIMSS döngüsünün hikâyesi

Şekil 5'te de görüldüğü gibi her uygulama döngüsünde uygulama öncesi ve sonrasında yapılan işlemler kendi içerisinde birbirini besleyen süreçler içermektedir. TIMSS'te tabakalı örnekleme yöntemi kullanılmaktadır: Türkiye'deki tüm okulların listesi uluslararası merkeze gönderilmekte bu merkez iki ayrı sınıf düzeyi için (4 ve 8. sınıf) önce okulları tesadüfen seçmektedir. Burada PISA'da da kullanılan İBB-düzey-1 bölgeleri kullanılmaktadır. Ardından seçilen okullar içerisinden tesadüfi olarak şubeler seçilmekte ve örneklem seçim süreci tamamen bağımsız ve yansız bir şekilde uluslararası merkez tarafından gerçekleştirilmektedir.

Uygulama sürecine ise öğrenciler matematik veya fen testi ile başlamaktadır. Sınav yaklaşık 36-45 dakika sürmekte, ardından 15-30 dakika arası mola verilmektedir. Diğer testin uygulaması tamamlandıktan sonra 5-15 dakikalık bir anket uygulaması gerçekleştirilmektedir. Sonuç olarak TIMSS asıl uygulaması yaklaşık yarım gün içerisinde tamamlanmaktadır.

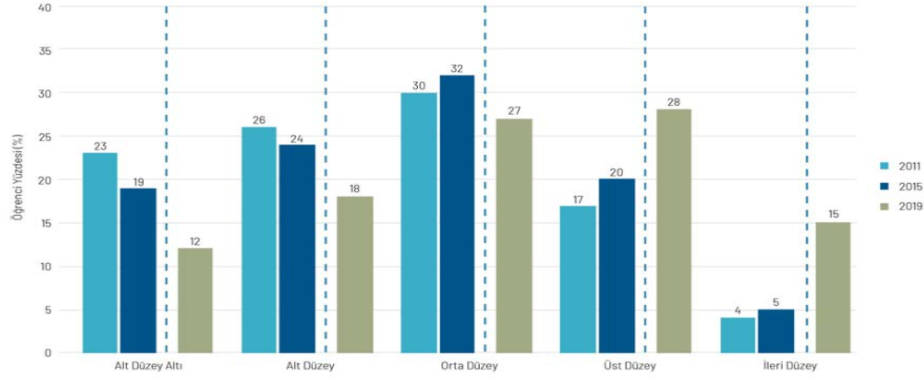
Millî Eğitim Bakanlığının 2017 yılında aldığı kararla Türkiye 2019 yılında ilk defa 4. sınıf düzeyindeki uygulamaya 5. sınıf örneklemi ile katılmıştır. Birçok katılımcı ülkede okul öncesi eğitim zorunlu eğitime dâhil olduğu ya da okul öncesi eğitim katılım oranları çok yüksek olduğu için genel ortalamaya göre ülkemizdeki 4. sınıf öğrencilerinin yaş ortalaması biraz düşük kalmaktaydı. Değişikliğin bir ikincisi sebebi ise oluşturulan değerlendirme çerçevesinin Türkiye'nin 5. sınıf eğitim programı ile daha uyumlu olmasıdır ve Türkiye diğer bazı ülkelerle beraber 5. sınıf örneklemi ile katılma kararı almıştır.

4. sınıf düzeyinde matematik alanında Türkiye 2011 yılında 469 puan olan ortalamasını 2015'te 483 puana ve son döngüde 523 puana çıkarmıştır. Türkiye 4. sınıf seviyesinde uygulamaya ilk defa 2011'de katılmaya başladığı için daha önceki döngülere dair bir sonuç bulunmamaktadır. 8. sınıf seviyesinde matematikte 429 puanla başlayan süreç, 496 puana kadar çıkmıştır. TIMSS'te 500 puan önemli bir ölçütü temsil etmekte olup bir uygulamadan diğerine aynı beceriyi ifade etmektedir. Türkiye ilk defa 4. sınıf seviyesinde 500 puanı aşmıştır.



4. Sınıf Düzeyi

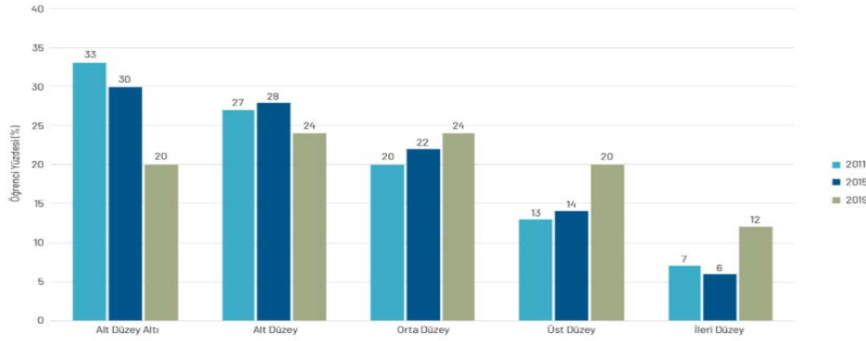
TÜRKİYE'DEKİ ÖĞRENCİLERİN SON TIMSS DÖNGÜLERİNDE
4. SINIF MATEMATİK YETERLİK DÜZEYLERİNDEKİ ORANLARI*



*Türkiye'nin katıldığı TIMSS döngülerindeki matematik yeterlik düzeylerine ulaşan öğrenci oranları. TIMSS 2019 döngüsüne Türkiye ilk kez beşinci sını düzeyinde katıldığı için Türkiye'nin 2019 ve öncesindeki döngülerindeki karşılaştırmalı performansı uluslararası raporda yer almamıştır.

8. Sınıf Düzeyi

TÜRKİYE'DEKİ ÖĞRENCİLERİN SON TIMSS DÖNGÜLERİNDE
8. SINIF MATEMATİK YETERLİK DÜZEYLERİNDEKİ ORANLARI*



*Türkiye'nin katıldığı TIMSS döngülerindeki matematik yeterlik düzeylerine ulaşan öğrenci oranları. TIMSS 2011 uygulamasında ortak maddelerin adaptasyon çalışması tekrar yapıldığı için Türkiye'nin 2011 yılı öncesinde katıldığı döngülerdeki sonuçları uluslararası raporda yer almamıştır.

Şekil 6. Matematik yeterlik düzeyleri oranları

Şekil 6'da görüldüğü gibi matematik yeterlik düzeyleri 4. sınıflar için incelendiğinde 2011 yılında öğrencilerimizin %4'ü, 2015'te %5'i ileri düzey yeterliğe sahipken son uygulamada bu oran %15'e çıkmıştır. 2011'de %7, 2015'te %6 olan ileri düzeydeki öğrenci oranlarımız ise son uygulamada %12'ye çıkmıştır. Alt düzey altı öğrencilerimizin oranı 4. sınıfta %23'ten %12'ye, 8. sınıf düzeyinde %33'ten %20'ye düşmüştür.

Fen bilimlerinde de çok benzer bir artışın olduğunu ve bu artışın özellikle 2019'da 4. sınıfta 463 puandan 526 puana ulaştığını ve puan artışının özellikle 2019 uygulamasında ivme kazandığı görülmektedir. 1999'da 433 ile başlayan ortalama puan, bugün 500 ortalama ölçek noktasını da geçip 515 puana ulaşmıştır. Bu anlamda PISA'ya benzer şekilde TIMSS sonuçlarımızda da hem 4 hem de 8. sınıfta matematik ve fen alanlarında önemli iyileşmeler söz konusudur.

Fen bilimleri alanında yeterlik dağılımları incelendiğinde ileri düzey seviyesindeki öğrenci oranları 2011'de %3'e, 2015'te %4'e ve son uygulamada %12'ye çıkmıştır. 8. sınıf



düzeyinde ise %8'den %13'e çıkmıştır. Alt düzeyin altında kalan öğrenci oranları ise 2014'te 4. sınıf düzeyinde %24'ten %18'e ve en son uygulamada %10'a kadar düşmüştür. 8. sınıfta ise 2011'de %21'den 2015'te %17'ye ve en son uygulamada da %12'ye düşmüştür. Bu sonuçlar, alt düzeylerdeki öğrenci oranlarımızın önemli ölçüde düştüğünü göstermesi açısından son derece önemlidir.PISA'da önemli ölçüde puanlarımız artmış olmakla birlikte OECD ortalamasına henüz ulaşamasak da TIMSS'in eğitim programıyla daha ilişkili olması sebebiyle müfredata dayalı değerlendirmelerde öğrencilerimiz daha yüksek performans göstermektedir. Bununla birlikte öğrencilerimizin daha üst düzey bilişsel becerilerde, iş gücü piyasasının talep ettiği daha karmaşık beceriler diyebileceğimiz becerilerde gelişim göstermeleri beklenen daha fazla alan bulunmaktadır.

PIRLS de IEA tarafından uygulanmakta olup okuma boyutunu içermesi sebebiyle TIMSS'i tamamlar nitelikte bir çalışmadır. PIRLS, öğrencilerimizin okuma becerilerine ve okuma düzeylerine dair önemli çıktılar sağlamaktadır. Türkiye PIRLS'e ilk defa 2001 yılında katılmış, 20 yıllık bir aradan sonra alınan kararla 2021 yılında tekrar dâhil olmuştur. Bu uygulamaya ilişkin sonuçlar 2022 yılının sonunda açıklanacaktır. Türkiye'nin PIRLS'e katılmasının bir diğer avantajı her iki çalışmaya katılan ülkelerin sonuçlarının (PIRLS ve TIMSS'ten elde edilen sonuçlar) ilişkilendirilebilmesidir.

4. OECD SOSYAL VE DUYGUSAL BECERİLER ARAŞTIRMASI

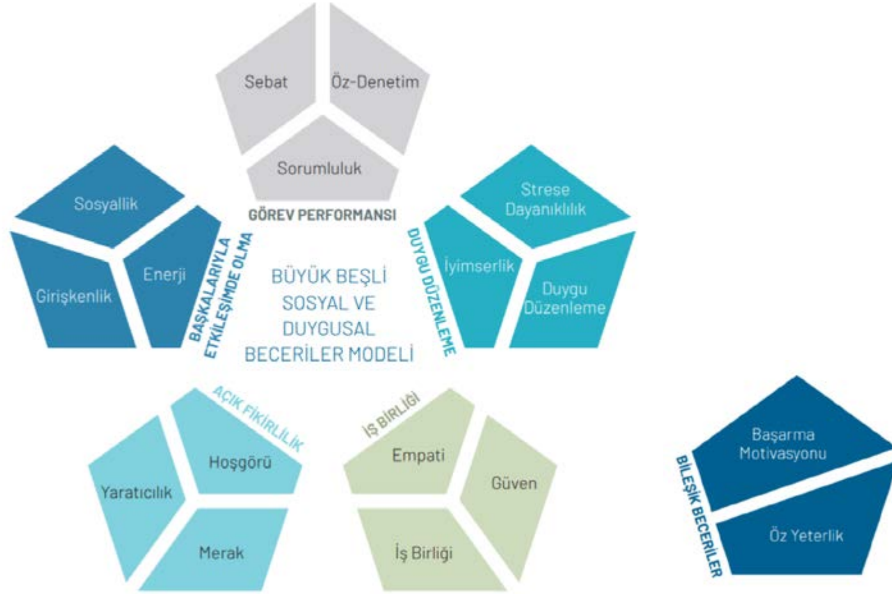
Dr. Hayri Eren SUNA

Bu ders kapsamında OECD tarafından geliştirilen ve Türkiye'nin de ilk döngüsüne katılım gösterdiği sosyal ve duygusal beceriler izleme araştırması tanıtılacak ve elde edilen genel bulgular paylaşılacaktır. Son kısımda ise Bakanlığımız tarafından yürütülen ulusal izleme araştırmaları hakkında bilgi verilecektir.

Yapılan eğitim izleme araştırmaları bağlamsal birçok bilgiye veri sunmaktadır. Sadece ülkelerin akademik performansını değil, aynı zamanda her bir ülkedeki öğrenci başarısıyla ilişkili birçok faktör hakkında bilgi sağlamaktadır. Öğrencilerin iyilik hâlleri, yaşam doyumu ya da çeşitli duyuşsal özellikleriyle diğer bir ifadeyle sosyal ve duygusal becerileriyle ilgili de birçok bilgi elde edilmektedir ama bu özellikler bu çalışmaların yalnızca bir boyutunu oluşturmaktadır. Tüm dünyada sosyal ve duygusal becerilerin önemi ile birlikte eğitim sistemlerindeki ağırlığının artmasıyla beraber 2017 yılında OECD tamamen sosyal ve duygusal beceriler odaklı yeni bir araştırma başlatmıştır. Türkiye de İstanbul'la bu çalışmayakatılmıştır.

Öncelikle bu çalışma neden önemlidir ve nasıl bir fark yaratmaktadır? Birinci olarak tümüyle sosyal ve duygusal beceriler dikkate alınıyor ki bu diğer tüm izleme araştırmalarından en önemli farkıdır. İkincisi 10 ve 15 yaş grupları ayrı ayrı ele alınmaktadır ki duyuşsal özellikler, sosyal ve duygusal özellikler yaşa bağlı olarak, özellikle de ergenlik dönemine bağlı olarak önemli değişimler gösterebilmektedir. Ergenlik sürecini de içerecek şekilde iki ayrı yaş düzeyinin dikkate alınması bu becerilerin nasıl değiştiği hakkında önemli çıkarımlar sağlamaktadır. Üçüncüsü ölçümler, dünyanın farklı bölgelerinde kabul gören beş faktörlü kurama dayalı olarak (Big five modeli) yapılmaktadır. Bu beş faktörlü model, Şekil 7'de verilmiştir.





Şekil 7. Büyük Beşli Sosyal ve Duygusal Beceriler Modeli

Beş faktörlü modeli oluşturan ana alanlar ve bu alanları oluşturan alanlar: açık fikirlilik (yaratıcılık, hoşgörü, merak), iş birliği (empati, iş birliği, güven), duygu düzenleme (iyimserlik, strese dayanıklılık, duygu düzenleme), görev performansı (sebat, öz denetim, sorumluluk) ve başkalarıyla etkileşimde olma (sosyallik, girişkenlik, enerji) olarak tanımlanmıştır. Her bir alanı oluşturan alanlar ilişkide ve temelde hepsi kendi içerisinde sosyal dayanağı olan sosyal beceri olarak tanımlanan özelliklerdir. Bir de birleşik beceriler diye tanımlanan ve bu beşliye dâhil olmayan ama onunla ilişkili olup sosyal beceriler açısından önemli görülen iki ayrı beceri ise başarma motivasyonu ve öz yeterlik algısıdır.

Bu çalışmayı özel kılan diğer bir özelliği de veri çeşitlemesinin (data triangulation) kullanılmasıdır. Veri çeşitlemesiyle kastedilen aynı sorular üzerinden hem öğrencilerden hem öğretmenlerinden hem de velilerinden bilgi toplanmasıdır ki bu tüm girdilerin/verilerin birleştirilerek sonuçların daha güvenilir hâle getirilmesine katkı sağlamaktadır. Veri çeşitlemesi kullanılmasının sebebi utangaçlık ya da kendine güven gibi duyuşsal becerilere ilişkin soruları, öğrencilerin oldukları gibi değil olmak istedikleri veya görünmek istedikleri şekilde cevaplandırma eğilimlerinin önüne geçebilmek; daha gerçekçi sonuçlara ulaşabilmektir. Bir başka deyişle soruları olduklarından daha öz güvenli veya daha az utangaçmış gibi kendilerini dışarı yansıtmak istedikleri şekilde cevaplama eğilimlerinin önüne geçilmeye çalışılmaktadır.

Bu çalışmaya katılan ülkeler ve şehirler şu şekildedir: Kanada (Ottawa), ABD (Houston), Kolombiya (Manizales ve Bogota), Portekiz (Sintra), Finlandiya (Helsinki), Türkiye (İstanbul), Rusya Federasyonu (Moskova), Güney Kore (Daegu) ve Çin Halk Cumhuriyeti (Suzhou). Sonuç olarak farklı ülkelerin katılımıyla ortaya çıkan kültürel çeşitlilik son derece önemlidir çünkü kültürel özelliklerle beş faktörlü modelde bahsi geçen sosyal ve duygusal özelliklerin yakından ilişkili olduğu bilinmektedir.



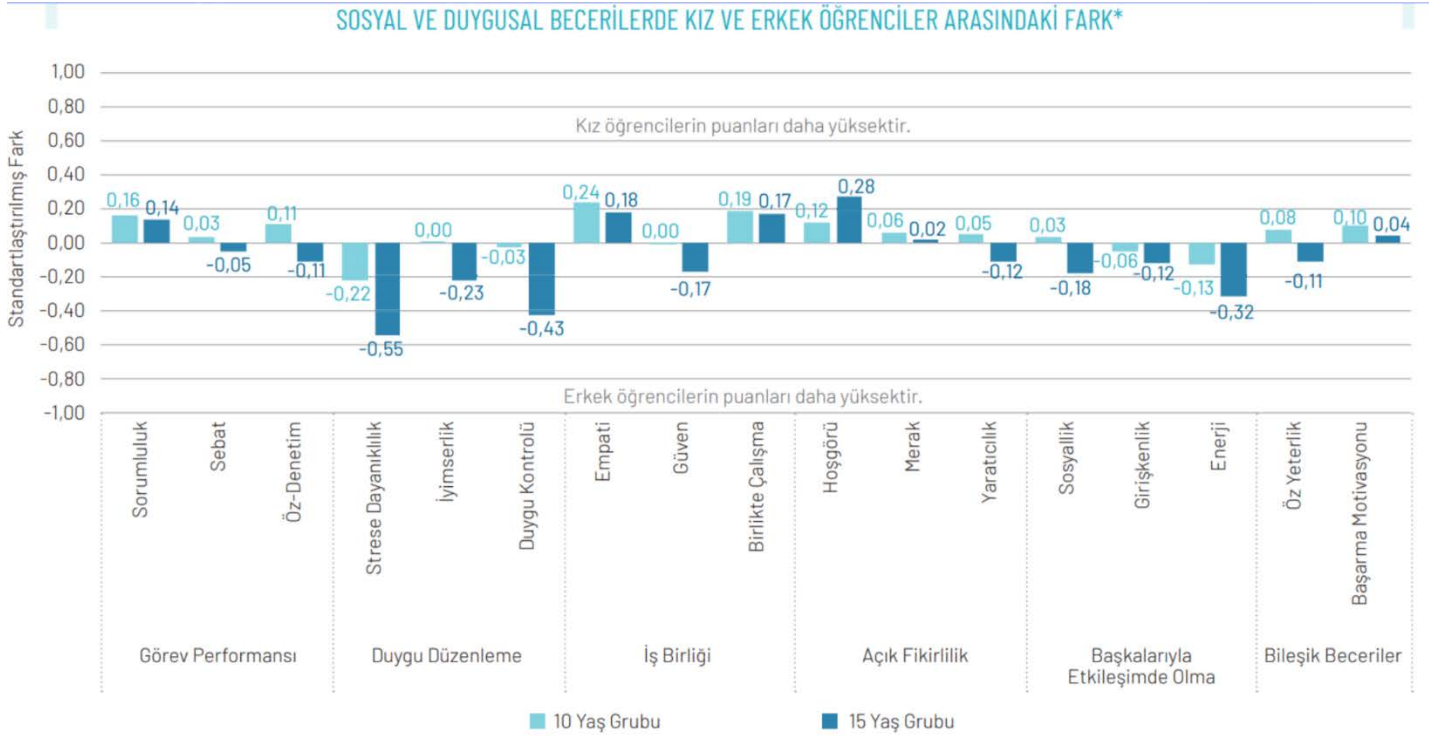
Çalışmaya Türkiye, 10 yaş grubundan yaklaşık 2.700 ve 15 yaş grubundan ise yaklaşık 3.200 öğrenci ile katılmıştır. Paylaşılan temel bulgular tüm katılımcılardan elde edilen genel bulgular olmakla birlikte Türkiye'nin bulguları genel bulgularla benzerlik göstermektedir.



Şekil 8. Yaş düzeylerine göre sosyal ve duygusal beceriler

Şekil 8'de görüldüğü gibi çalışmanın enteresan sonuçlarından biri 10 yaş grubunun sosyal ve duygusal beceri puanlarının 15 yaş grubundan daha yüksek olmasıdır. Şekilde üst tarafa doğru olan çizgiler 15 yaş grubunun puanlarının yüksek olduğunu ifade ederken aşağıya doğru giden çizgiler 10 yaş grubunun sosyal ve duygusal beceri puanlarının daha yüksek olduğunu ifade etmektedir. Buna göre neredeyse tüm becerilerde 10 yaş grubu 15 yaş grubundan daha yüksek puanlara sahiptir. 15 yaş grubunda ergenliğin etkilerinin yoğun görüldüğü düşünüldüğünde sosyal ve duygusal beceriler üzerinde 10 yaştan 15 yaşa geçerken önemli bir düşüş olduğu aşikârdır.

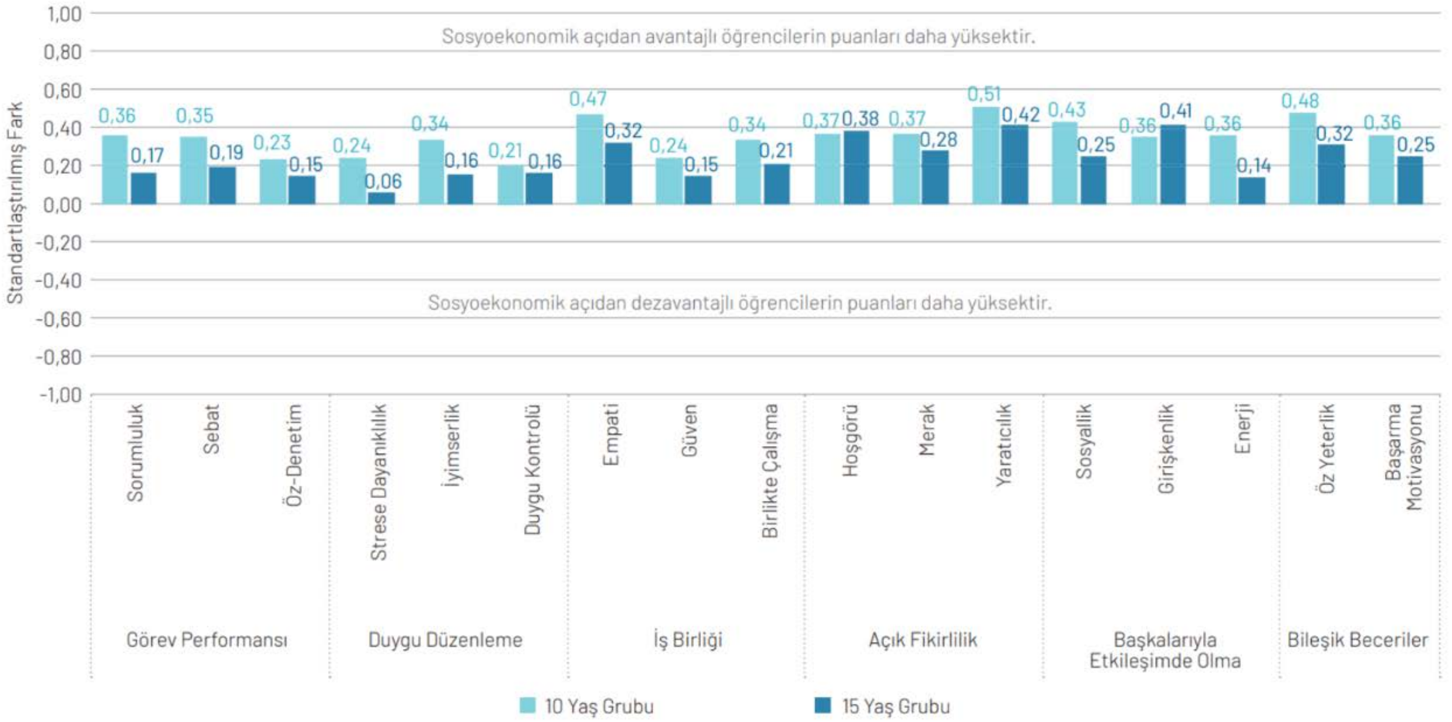




Şekil 9. Cinsiyete göre sosyal ve duygusal beceriler

Şekil 9'da görüldüğü gibi cinsiyete göre bu sosyal ve duygusal becerilerin nasıl değiştiği incelendiğinde sosyal ve duygusal becerilerin önemli farklılıklar gösterdiği gözlenmiştir. Strese dayanıklılık, duygu kontrolü ve enerji gibi bazı özelliklerde erkek öğrencilerin puanları yüksekken sorumluluk, empati, birlikte çalışma, hoşgörü ve merak gibi becerilerde kız öğrencilerin puanları daha yüksektir. Açık mavi ile gösterilen çizgiler 10 yaşı, daha koyu renktekiler ise 15 yaşı ifade etmektedir. 10 yaştan 15 yaşa geçişte yön değiştiren bazı özellikler bulunmaktadır. Örneğin öz denetim puanının 10 yaş grubundaki kız öğrencilerde, 15 yaş grubunda ise erkek öğrencilerde daha yüksek olduğu görülmektedir. Benzer durum öz yeterlik puanı için de geçerlidir. Başarma motivasyonu ise her iki yaş düzeyinde de kız öğrencilerin daha yüksek puana sahip olduğu bir özelliktir. Özetle çıkarmamız gereken en önemli sonuç, sosyal ve duygusal beceriler anlamında kız ve erkek öğrencilerin önemli farklılıklara sahip olduğu ve bu becerileri geliştirmeye dönük her türlü müdahalede cinsiyetler arası farkların dikkate alınması gerekliliğidir.





Şekil 10. Sosyoekonomik düzeye göre sosyal ve duygusal beceriler

Önceki çalışmalardan sosyoekonomik düzeyin öğrenciler açısından son derece belirleyici olduğunu bilinmektedir. Sosyoekonomik düzey, ailenin sahip olduğu eğitim düzeyi ve gelirden öğrencinin evde sahip olduğu eğitim imkânlarına, okuma imkânlarına hatta fiziksel imkânlara (keni odası olup olmaması, internet bağlantısına sahip olup olmaması) kadar her türlü eğitim ve imkâna dayalı fırsatı kendi içinde barındıran yapı olarak tanımlanmaktadır. Sosyoekonomik düzeyi yüksek olan öğrencilerin hem toplumda hem de eğitimde birçok konuda daha avantajlı olduğu bilinen bir gerçektir. Şekil 10'da da görüldüğü gibi sosyoekonomik açıdan avantajlı öğrencilerin daha yüksek sosyal ve duygusal beceri puanlarına sahip olduğu gözlenmiştir. Ölçülen bütün özelliklerde sosyoekonomik açıdan avantajlı öğrenciler daha yüksek puanlar almıştır. Özellikle yaratıcılık, empati, öz yeterlik gibi becerilerde puanları daha yüksektir. Bu durum sosyoekonomik açıdan dezanavantajlı öğrencilerin sosyal ve duygusal açıdan daha fazla desteklenmesi gerektiğini açıkça ortaya koymaktadır.

Sosyal ve duygusal beceriler bağlamında bireyin öz yeterliğini geliştirmek, her türlü konuda onu daha yeterli hissettirmek veya kendine yönelik yeterlik inancını artırmak önemlidir ama bunlar akademik başarı ile ne kadar ilişkilidir? Burada iki özellik ön plana çıkıyor: Birinci özellik olan merak, hem okuma hem de matematik becerileriyle anlamlı bir ilişkiye sahiptir. İkinicisi de sebat, yani aynı konu üzerinde öğrenene kadar sabırla çalışmaktır. Bu iki özellik 10 yaş grubunda akademik başarıyla doğrudan ilişkili görünmektedir. 15 yaş grubu incelenecek olursa yine merak ve sebata ek olarak kendini güvende hissetme, sorumluluk bilincine sahip olma ve yaratıcılık özelliği de farklı başarı göstergeleri ile ilişki göstermektedir. Bu becerilerin desteklenmesi akademik başarıyı da yükseltme potansiyeline sahiptir.

Öğrencileri desteklemek adına bu becerilerini geliştirmek için neler yapılabilir? Çalışmalar, spor etkinliklerine katılımın yaratıcılık ve merak puanlarını önemli ölçüde artırdığını göstermektedir. Örneğin müfredat dışı spor aktivitelerine katılan öğrencilerin katılmayanlara göre yaratıcılık puanları %17 ile %16 arasında daha yüksek, keza merak puanları da %10 ile



%15 arasında yaş düzeyine bağlı olarak daha yüksektir. İstanbul'da da bu oranların uluslarası ortalamadan daha yüksek olduğu gözlenmiştir. Sanat etkinlikleri de öğrencilerin merak ve yaratıcılık becerilerine önemli katkı sağlayan etkinliklerdir. Sanat etkinliklerine katılan öğrencilerin yaratıcılık puanları %23 ile %30 arasında, merak puanları ise yaklaşık %15 daha yüksektir. Müfredat dışı sanat ve spor etkinliklerini desteklemek bu öğrencilerin akademik becerileri ile ilişkili olan sosyal ve duygusal becerilerini artırmaktadır.

Okul iklimi ile ilgili çıktılar ise rekabetçi okul iklimi, veli beklentileri ve öğretmen beklentilerini kapsamaktadır. Akademik çalışmalardan, öğrencilerden beklenti arttıkça (beklentinin gerçekçi olması koşuluyla) öğrencilerin akademik performansları ve kendi öz yeterlik algılarının arttığı bilinmektedir. Veli beklentileri ve öğretmen beklentileri arttıkça öğrencinin psikolojik iyilik hâli ve sınav kaygısı da artmakta ancak bir taraftan da psikolojik kaygıları ve yaşam doyumları kısmen artmaktadır. Diğer taraftan rekabetçi okul iklimi ve okul içi rekabeti destekleyen politikalar arttıkça öğrencilerin özellikle psikolojik iyilik hâlinde önemli eksikler olduğu görülmektedir. İstanbul'un yaşam doyumu açısından 10 yaş grubunda uluslarararsı ortalamaya oldukça yakın ama 15 yaş grubunda uluslararası ortalamanın altında olduğu gözlenmiştir. Tüm katılımcı şehirlerde 10 yaştan 15 yaşa gelindiğinde öğrencilerin yaşam doyumu azalmaktadır. Bu nedenle öğrencilere yaşam doyumu sağlayacak, onları kabul edecek olumlu okul iklimi yaratmak son derece önemlidir. Okullarımızda zorbalığa maruz kalma, büyük oranda öğretmen ve öğrenci ilişkisi bağlamında çözülebilmektedir. Öğretmen ve öğrenci ilişkisi arttıkça, daha etkileşimli hâle geldikçe zorbalığa maruz kalma durumu azalmaktadır. Öğretmen ve öğrenci ilişkisi arttıkça okula aidiyet de artmaktadır. İstanbul, katılımcı şehirler arasında öğrencileri okula aidiyet oranı yüksek şehirlerden biridir.

Ulusal izleme çalışmaları kapsamında Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) araştırması, üst düzey bilişsel becerilere öğrencilerimizin ne kadar sahip olduklarını incelenmekte; Türkçe-Matematik-Fen Öğrenci Başarı İzleme Araştırması (TMF-ÖBA) ise öğrencilerimizin eğitim programına bağlı kazanımlara ulaşma düzeylerini belilemek üzere yapılmaktadır. Son olarak 2019 yılında ilk defa tamamen bilgisayar ortamında gerçekleştirilen okuma, dinleme, yazma ve konuşma olmak üzere dört temel becerinin hepsini kapsayan Türkçe dil becerilerinin izlenmesi araştırması yapılmış ve yaygınlaştırılmaktadır.

Bu eğitimde kullanılan tüm görsel, bilgi ve yorumlara detaylı şekilde Eğitim Analiz ve Değerlendirme Raporları (https://raporlar.meb.gov.tr) aracılığıyla ulaşabilirsiniz.